Premium Home Theater

Design & Construction

Premium Home Theater

Design & Construction

Earl Geddes
with
Lidia Lee

GedLee LLC
Novi, Michigan

Book cover design by: Mark W. Geddes
Editing: Ron Barwinski

Printed in Hong Kong

ISBN 0-9722085-1-8

For Nathan and Abbey

with special thanks to:
Mike Maloney,
Duke LeJeune, and
Chris Collins
for helpful suggestions and
proof reading

CONTENTS

CHAPTER 5

CHAPTER 6

CHAPTER 7

Introduction

It is always important to define the intent of a book and sometimes what it is not. Basically this is a technical book, one where the primary topics will be the technical aspects of Home Theater (HT) design. It is aimed at a reader who has some knowledge of audio-visual systems, but is not necessarily skilled in either. The use of math is limited, but the topic of HT cannot be fully understood without a minimum level of technical capability. To do so would be analogous to attempting accounting without using arithmetic or law without using logic (although it sometimes seems that this is common). My intention is to arm the reader with sufficient background in the pertinent subject areas so that they could design and build, or have built, a HT of the highest caliber.

I will also limit my discussion to what I will call high "value" designs and components. One can spend almost an unlimited amount of money on a HT, but that is not my interest here. My intent is to describe the most cost effective—the highest value—designs and components for the reader who has a limited budget and I doubt that anyone reading this text would not.

In my own HTs, I have done all of the construction work and this practical experience coupled with my technical knowledge of the subject matter has given me a unique insight into the design of HTs. I do not necessarily advocate a do it yourself approach to construction. This approach is by far the most cost effective, but it does require some very specialized skills and equipment and a huge amount of time. It can be done if one has the time and inclination, and, most importantly, the knowledge contained in this text, but, by far, the most common approach will be to hire someone else to do the heavy work. In either case, the knowledge gained from reading this text will prove to be the difference between success and failure.

This book would be great background for any HT installer. With it, you will gain a fundamental understanding of what needs to be done. I strongly recommend its use by this group. I would also recommend it to the consumer who wants to direct or at least understand the options available to them in a HT installation. Hiring a “pro” installer and giving them a free hand in the design and implementation may not result in the most cost effective or highest performance system. With the information presented in this book, it is hoped that the installer will get better and the consumer wiser. In that way, all parties involved will be more satisfied with the process and the final results.

If you are considering doing the work yourself, I do not discourage that approach. I have known people to build superior rooms by themselves, on their first try, with little prior knowledge or experience (but with the same kind of guidance that this book will give) and I document one such case in this text. But, to me, the ideal situation is for the consumer to do the basic floor plan selection and room design along with the specification of the system components then hire a contractor to implement it. With this approach a great deal of “negotiation” has to take place at the beginning to find a *workable* compromise between the owners desires and the installers’ capability.

Many of the construction techniques proposed in this text were developed through my own experience with construction and are not all that common. Some techniques may not be acceptable to every building inspector throughout the world or, most importantly, may not meet the numerous local building rules and requirements (code). One should always check the local codes to see if the construction meets them and modify the construction as appropriate. My techniques may also be somewhat difficult to implement and, in my experience, contractors always prefer to steer clear of this kind of unique work. If one allows the contractor to utilize those components and techniques that they are familiar with, then the quote may be better, but, unfortunately, these components and techniques may not always be the best value or performance. As a result, it has been my experience that the more common components used for HT tend to be premium priced ones (not always a good value), and the selection of alternates can often lead to equivalent performance at a substantially reduced price or better performance at the same price. When both parties are knowledgeable about the technical background, the best compromise between what the client desires and what the contractor prefers can usually be reached. Without this mutual knowledge, the compromises often lead to higher costs or disappointing results, and make no mistake about it, these costs (and sometimes the savings) can be substantial.

I once saw a television show on “home theater” which estimated the costs at $60,000–$100,000. My rooms typically run between $15,000–$25,000, excluding labor, and for a significantly better design. Based on this price point, my approach would cost less than half of what the television program claimed the costs would be. The potential savings, as well as the performance enhancement, more than warrant the reading of this text if one is price conscious—at these costs who wouldn’t be?

The phase at which the consumer has the most influence on the success of the final project is the earliest stage of design. This is where the most important and wide-ranging decisions are made. I find that this is often where the weakest aspect of many of HTs originate. Many designs have significant flaws in their ability to isolate the sound from within and without. They also suffer from a poor selection of components—too much money spent in one area and not enough in another. Both of these aspects must be very early considerations in the design. It is not uncommon to run out of money at the end of the project, just when one is about to buy the speakers, and then to short change this extremely important consideration and purchase.

Many HTs look better than they perform which, I believe, is because most of the attention has been paid to making the room look good (typically by making it look like a commercial theater) without nearly as much attention being paid to its optical and acoustical performance. I will stress throughout this text that looking like a commercial theater is not really what one wants in a HT. If cost and/or performance are high priorities, then the room will almost certainly not look like a commercial theater. My clients have sometimes been leery of the early suggestions for the aesthetic features of the space. For instance, when I propose creating dark windowless rooms with masonry walls, kind of like a dungeon, clients (particularly the wives) often shudder. The fear is that the HT will lack any aesthetic appeal. The fact is that my rooms look different, but my clients have always found them to be ascetically appealing in the final analysis, particularly after they audition the superior performance of my HT designs. Fashionable HT magazines clearly put a very high priority on appearance, doing things and using materials and furnishings that I avoid as both costly and ineffective. They clearly are not using the materials and construction techniques that I recommend.

Like anyone else, I have my own personal preferences in aesthetics and interior decorating and there is no doubt that my designs lean toward those preferences—how could they not? I like simple styling with liberal use of wood for finish and trim; subdued colors (for which there is a good technical justification); and I have a strong preference for leather upholstery, which is not only practical and attractive but has very desirable acoustical properties. I would like to make the point here that after reading this book you should have sufficient knowledge to be able to select design themes, materials and colors which are compatible with both to your own personal tastes and which work well technically. Without understanding the technical background presented in these pages, it is difficult to mix subjective tastes with performance requirements in a way that will not compromise the end product.

I encourage you to read Appendix I on Subjectivism as I think that the reader would benefit from an understanding of my views on this subject.

Finally, a word on references. References have not been supplied in this book except for a single instance of my own text. This is because today, most cross referencing is done via a web search. I obtained much of the information in this book this way. I could have put in the links to these references, but I have found that links change so often that it is better just to give a list of keywords. I have provided such a list in the form of side notes for such words. The first place to start any investigation into more details is to use these keywords.

CHAPTER 1 SIGNALS AND SYSTEMS

System
composite makeup of components to perform a particular function.

Signal
a set of data representing some underlying meaning.

Home Theater (HT) is founded in two different but surprisingly similar technical areas—audio and video (AV). Both of these technologies have a common underlying technical basis and because these technical concepts are common to both areas, I have placed this material first. Therefore, I am going to start this text with the toughest material. I do this not because I enjoy torturing the reader from the outset, but because this material is fundamental and common to all the technology that makes up HT. Both audio and video are **systems** that manipulate **signals**. To be sure, the signals and systems in each of these areas are different, but the underlying concepts for dealing with them are virtually identical.

This means that I can cover the fundamentals of the two subjects with a single technical discussion—a tremendous savings of effort. The unfortunate part is that I must discuss these concepts generically, without specific reference to their applications since they can be applied in different situations in different ways. Admittedly, this creates a double-edged sword of conciseness versus elucidation. I hope that the reader will make the effort to follow the discussions since doing so will pay big dividends in later chapters when more specific applications are discussed.

1.1 Signals

Value
a numeric attribute of a physical quantity denoting its amount.

Signals are present all around us. A signal is any variation in a **value** (a numeric attribute of some physical quantity that denotes its amount, such as volts or pressure) over time which contains useful information. Voice is a signal, as is a doorbell, as is video, etc. Noise is not generally thought of as containing information, although there can be exceptions (such as when your car makes a "noise" that signals something is wrong). Thus, there are basically two classifications of time variations—signals and noise. One contains information—usually the desired information—and the other does not. Note that what constitutes a signal and what constitutes noise does not have an absolute distinction—one person's signal can be another's noise.

The concept of a signal is useful since it allows us to have a discussion of a data stream without the need to reference the underlying physical system that carries this data stream. For instance, we can talk about a signal and its characteristics without having to reference the actual physical quantity that we are talking about, such as the pressure or volts, etc. or even the intended end product, i.e. sound or video. We can let these signals flow from one system to the next with little concern about the details of the underlying physical system they are currently in. For example, when someone is talking on the phone, the acoustic voice signal is converted into an electrical current signal which travels along cables. This is often done as an optical digital signal. The signal then arrives at its destination ready to be converted back along an identical but reverse process into a sound signal at the receiver—human hearing. At every location the "signal" should be the same, but it will take on different physical forms (electricity, light, sound, even analog or digital) at every stage.

We hope, or assume, that the signal content remains constant along the entire path, but it never does. I will talk in later chapters about how a signal can be changed as it progresses along a transmission path, and I will show how the modifications to these signals are created and defined. How these modifications can change the human perception of these signals and the important question of when these modifications actually cause a change in the perception or not will also be discussed. Since all signals are changed by the systems that propagate them, knowing what changes affect human perception and what changes do not is a crucial distinction. I will attempt to define those signal changes which have been shown to be the principal contributors to deviations in perception. However, I will not be able to discuss all of the contributors down to the most inconsequential ones. The result of this limitation will be that I may not discuss aspects of systems and signals which are felt by the reader to be important. That does not mean that I don't agree to their existence or perhaps their importance, only that I may not agree with their priority.

1.1.a Signal Level

Scale
applying a number to the data value.

Even though a signal can exist without a reference to a specific physical quantity, we will need a way to define its **value** at any given moment in time, i.e. a way to **scale** it. Scaling is straightforward if we simply use the definition of the physical quantity carrying the signal, voltage in Volts or sound pressure in Pascals for instance, but it will turn out that another scaling method is far more convenient for AV signals. Since the concept of a signal usually implies that it has information content, it also implies a human perceptual interpretation. When human perception is involved a different scaling definition is usually desirable.

The reason for this different scaling is because human perception mechanisms for signals in nature, light, sound, and even touch or smell to a lesser extent, tend to respond equally to ratios of excitation rather than the actual level of excitation. For example, subjects will judge each doubling of the sound pressure as being a perceptually equal increment. This means that going from 1 unit to 2 units is perceived as the same perceptual change as going from 10 units to 20 units, even though the actual physical increments are one and ten respectively. There are biological reasons why humans react in this way, but that is a topic beyond the scope of my intended subject. Relationships which have this characteristic are defined mathematically as logarithmic. A more detailed description of the **logarithm** can be found in Appendix II and this appendix also discusses some important characteristics of the dB scale, which I am going to introduce next.

Logarithm
a mathematical relationship wherein equal percentage changes have equal numbers.

A method for scaling a logarithmic relationship, which has found almost universal application, is called the deci-Bel (dB, a tenth of a Bel, after Alexander Graham Bell). The dB is the most common unit of measure in both audio and video. The dB scale gives a signal scaling which is more in line with human perception. It is also a scaling method which has a number of useful features, most of which are discussed in the appendix.

1.1.b The Time-Frequency Relationship.

Frequency
the number of cycle, per second, that a waveform exhibits.

Pitch
the perceived tone of a sine waveform.

One of the most important concepts in any discussion of signals is the relationship between the time and frequency domains. (In video, it is the space and frequency domains that we are interested in, but I think that the time domain is more familiar so I will focus on that one for the moment.) The time domain is something that we all have a basic understanding of even though it is hard to actually define. We typically have a good conceptual understanding of **frequency** because our daily lives contain all kinds of sounds and we can readily distinguish a high frequency sound from a low frequency one. The problem with this description is that what we perceive as frequency is actually called **pitch**, which is different from frequency. I don't want to get too deep into this difference other than to note that our common daily experience with "frequency" is really a perceptual experience with pitch. Once again it is basically a logarithmic relationship that connects the two.

Amplitude
the value of a waveform at some point in time.

Hertz
the unit of frequency—one cycle per second.

Sine
a convenient trigonometric function for defining the simplest of waveforms
(See App. III).

Frequency is defined as the number of repetitions that the signal level undergoes in a given period of time. Figure 1-1 shows three **sine** waves of different frequencies and **amplitude** in the time domain, that is, time as the horizontal (x) axis. These waves have frequencies of 1 cycle per second or **Hz** (**Hertz** after the German physicist Henrich Hertz), 1.5 Hz and 4 Hz. If

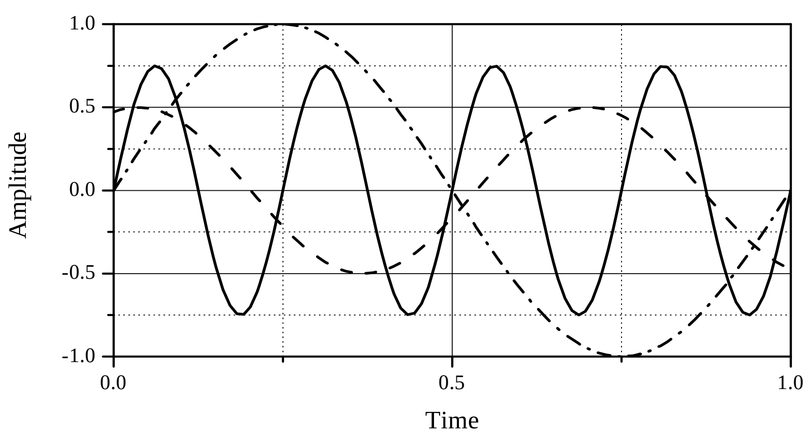


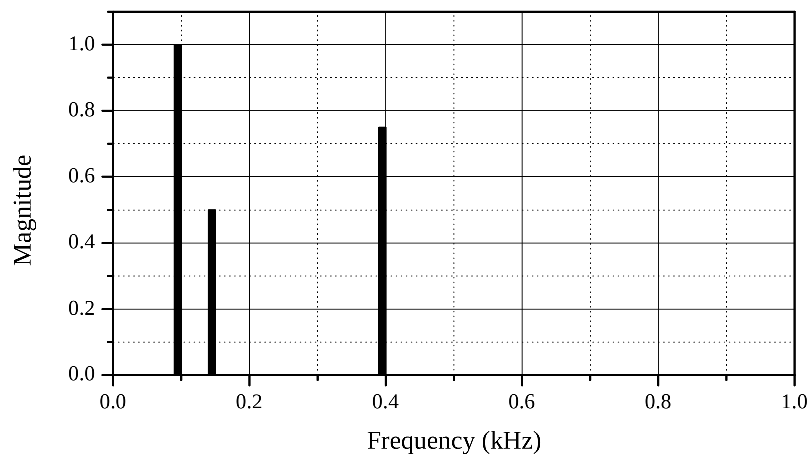
Figure 1-1.
Three sine waves of different amplitude, phase and frequency.

the time axis were in milliseconds (ms, 1/1000 of a second) instead of seconds then these waves would be 1000 Hz (or 1 kHz), 1.5 kHz and 4 kHz respectively. These higher frequencies are more common to us than the very low ones, so I prefer to show time in milliseconds since this time scale lends itself better to a more common every day experience.

Frequency domain
the technique of studying signals and systems by looking at their frequency characteristics.

Consider an alternate way of showing the data in Figure 1-1. Let me call this new way of looking at the data the **frequency domain**. In Figure 1-1, each of these waveforms can be described by an amplitude and a frequency—two simple numbers. If I now plot these numbers on a new graph with amplitude as one axis and frequency as another, then I will get a plot as shown in Figure 1-2—which is called the frequency domain. Note that time is no longer directly apparent in the frequency domain just as frequency is not directly apparent in Figure 1-1. The two plots do, however, contain exactly the same information.

Figure 1-2. Frequency view of the three waves shown in Figure 1



Magnitude
the peak value attained by a sine wave independent of when it occurs.

RMS
the square root of the mean (or average) value of a waveform.

In the frequency domain, each waveform is located at its specific frequency and is drawn as a line whose height indicates the waves amplitude. The vertical axis is labeled as the **magnitude**, which is the peak value of the waveform independent of the point in time at which this occurs. The magnitude and the amplitude are slightly different things where the term amplitude is usually used to refer to the value of the waveform at any given time in the time domain and the term magnitude is used to refer to the waveforms maximum value. Sometimes, we might see magnitudes given as **RMS values** (Root Mean Squared). RMS is basically an effective value, a sort of average value for the waveform. The term magnitude is usually used to refer to frequency domain levels.

Phase
the starting location of a sine wave relative to its zero crossing or another sine wave.

In Figure 1-2, the 1.5kHz wave does not start at zero as the other two do, this wave has been shifted along the time axis. The starting point of the waveform is called its **phase**, completely analogous to the “phase of the moon”. The time delay that causes a given phase depends on the frequency of the waveform. The important thing to note is that to completely describe the time domain data in the frequency domain, I also need to know its phase. Thus in the frequency domain the waveform is represented by its frequency and its magnitude and phase or equivalently its **complex magnitude**. It is complex because the magnitude can be described by two numbers known as the real and imaginary parts of a complex number.

Complex magnitude
the magnitude and phase of a waveform express as a two part—complex—number.

It’s not really too important to note this “complex” aspect of the magnitude, but it is described in more detail in Appendix III - Complex numbers. The reason that I even acknowledge it is because mathematically the calculations are all done in complex arithmetic. Fortunately, I will almost never need to resort to this complication. I will plot magnitudes as a single real number and sometimes show the phase—the phase being of lessor (but not insignificant) importance. If it is desirable to show the phase, then it is usually done as a second plot or a second line on the same plot but with a different scaling. Note that the magnitudes as shown in Figure 1-2 are blind to the waveforms starting point, i.e. the phase.

Complex waveforms
waveforms composed of a multitude of sine waves of different frequencies.

The relationship between the time and frequency aspects of a sine wave can be extended to **complex waveforms**. Consider the waveform shown in the top half of Figure 1-3. This waveform might be exhibited, for example, by a musical instrument since it is periodic. With a period of 1 ms it would have a base frequency of 1000Hz, or 1kHz. The period of a repetitive signal is the length of time that passes before the signal exactly repeats itself. By using a mathematical technique called the **Fourier Series**, I can decompose the complex upper waveform into a set of pure sine waves. I have shown this decomposition in the bottom half of the figure. This series—it is called a series because it is the sum of a series of independent waves—is named after the French mathematician “Fourier” who first studied its properties. In the case shown here, all of the sine wave are related by integral frequencies. The longest waveform in the series is called the **fundamental** and the higher frequency (shorter) ones are called **harmonics**. The integer relationship is important because it means that only waves which are *n* times the fundamental, where *n* is an integer, are allowed in this series. This requirement is a direct result of the fact that the Fourier Series repeats itself, it is always periodic. Its values continuously repeat in any time interval that is outside of the time **period** of the fundamental—1.0 ms in Figure 1-3. Any signal represented by a Fourier Series must repeat itself on exactly this period and so only waves which have frequencies that are integer multiples of the lowest frequency and, hence, synchronous to it are allowed.

Period
the time for a periodic waveform to repeat itself.

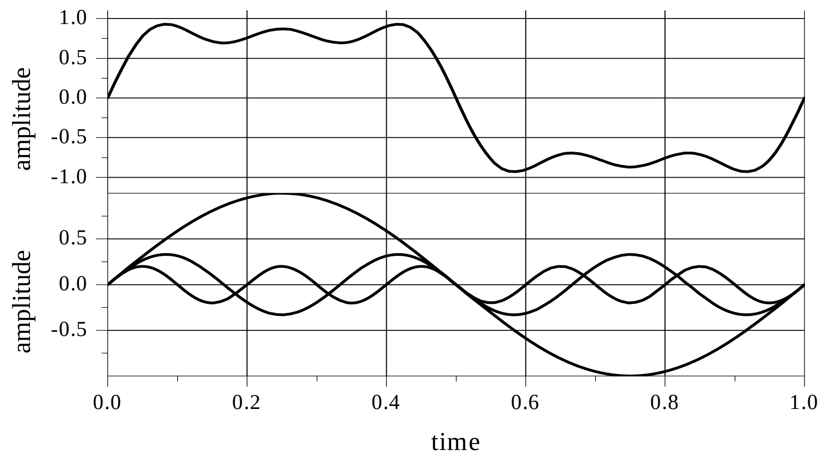


Figure 1-3.
A complex waveform and its constituent parts.

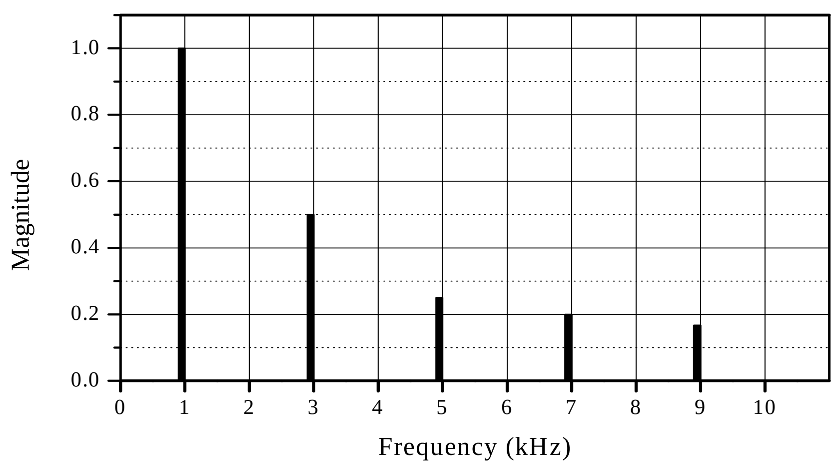


Figure 1-4.
Frequency domain representation of a square wave.

Figure 1-3 also has a frequency domain representation as shown in Figure 1-4. In this figure, I have plotted the components by their order (the integer of their multiple) which is of course proportional to their frequency. There is no need to plot the phase of the components since it is obvious that they are all in phase. I have shown the first to the fifth order for the Fourier Series decomposition of a square wave. Figure 1-3 only shows the first three components of this series, but it should be clear that adding more terms would lead to a more squared off waveform.

The reader should take note of the close relationship between the discussion of Figure 1-1 and Figure 1-2 and the Fourier Series component representation shown in Figure 1-3 and Figure 1-4. They are basically the same thing. The Fourier Series shows us how to map signals from the time domain to the frequency domain in a concise mathematical framework.

The Fourier Series is ideal for decomposing harmonic waveforms as would appear from a single instrument playing a continuous periodic tone, but this is hardly the most general form of a signal. The world is composed of harmonic, in-harmonic and transient signals as well as music, which is made up of a multitude of individual instruments with all of these signals being present simultaneously. In-harmonic signals come from instruments like cymbals and drums which have waveforms which do not have an integral relationship between their components and almost all musical instruments that have both transient and steady state signal components. In order to be able to decompose a completely general (real life) waveform, I will need to extend the Fourier Series concept to a close relative which allows signals which are not necessarily periodic.

Window
the limited time frame in which we look at a waveform.

Consider an impulse waveform as I have shown in Figure 1-5. For now, let's simply ignore what this waveform looks like outside of the "**window**" that we are currently looking at ($0 - .3s$) and go ahead and find its Fourier series components. The result of this exercise is shown in the top of Figure 1-6. Note that the components are still all harmonics, but of the lowest frequency component, defined by the window, at $4Hz$ The principle waveform period is seen to be at 24 Hz (the peak value) which corresponds to a period of about .04 s.

Consider the **time base** of Figure 1-5, the length of time shown in the plot, the window, which I will also assume is the length of time over which I take the data. In this example it is $.3s$. If I let the time base become much

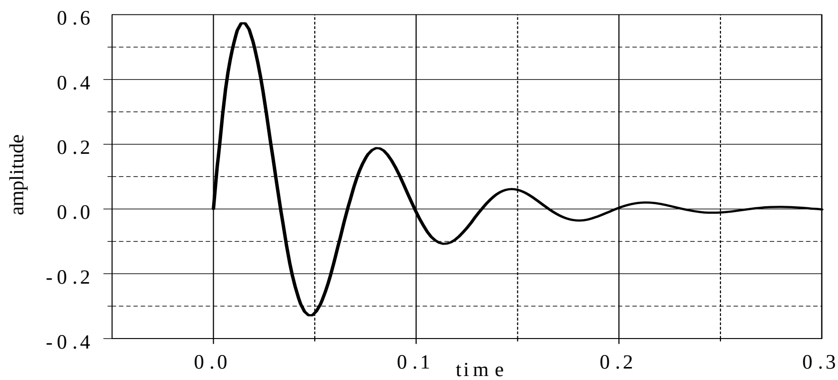


*Figure 1-5.
A non steady state
signal—an impulse.*

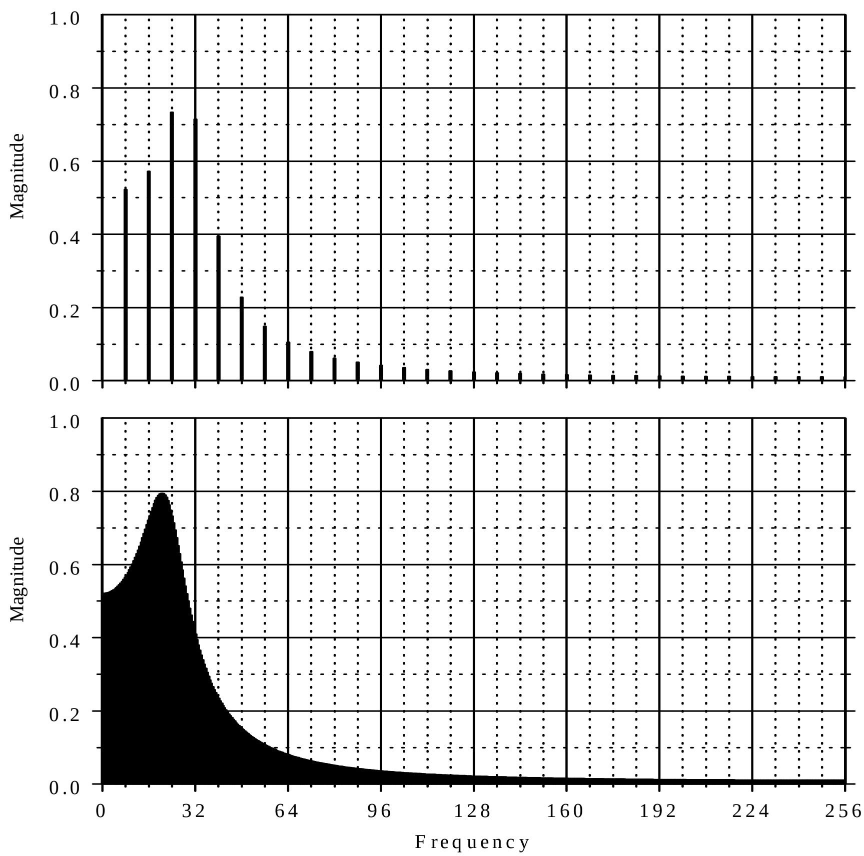


*Figure 1-6.
The Fourier Series
terms in the expansion
of Figure 1-5 for two
different time windows.*

larger, say 8 times longer or about 2.5s, then I will get the results shown on the bottom of Figure 1-6. Both curves are plotted with discrete lines to represent the frequency components, but in the lower plot these lines have become very dense. The lines are still all harmonics of the lowest period, but this period has become much longer and the lowest frequency is now $4/8 = .5$Hz. The peak value is still at 24Hz however. The results have not changed, only the resolution of those results has increased.

If I were to let the period in Figure 1-5 go to infinity, then the discreet line structure of the representations in the above figures would become infinitely dense and would create what is called a continuum—a continuous curve not discreet lines. In this later case we usually drop the filled-in area under the curve and draw a line from point to point. The continuous version of the Fourier Series—the one where the time base goes to infinity—is called the **Fourier Transform**. It is called a transform since it transforms data from one continuous domain—time—into another continuous domain—frequency. For our immediate purposes, we will limit our discussions to transformations between the time domain and the frequency domain although the Fourier Transform (as I will show later when I talk about video) has a much broader applicability and holds for transformations between many different kinds of variables (domains) and also in multiple dimensions such as space.

Fourier Transform
the continuous frequency version of the Fourier series where the time window becomes very long.

Applying the Fourier Transform to a time domain signal results in its frequency **spectrum**—basically a frequency domain plot like that shown in Figure 1-6. A spectrum is defined by two values at every frequency. The two values can be given in two equivalent sets, as **real** and **imaginary** values (a **complex number**) or the far more common terms, magnitude and phase. I discuss these quantities and the relationship between them in Appendix III - Complex numbers.

Spectrum
the continuous frequency representation of a signal.

In audio it is common to view the spectrum in dB with a logarithmic frequency scale (making it a log-log plot). This plotting standard is useful because the picture created closely represents what we would actually hear (perceive)—the pitch being approximately logarithmic and the level scaling in dB. Appendix II - Logarithms shows some of the important relationships that apply in the log-log domain. Exponential relationships appear as simple straight lines in this type of plot which can be a significant simplification in visualization and calculations.

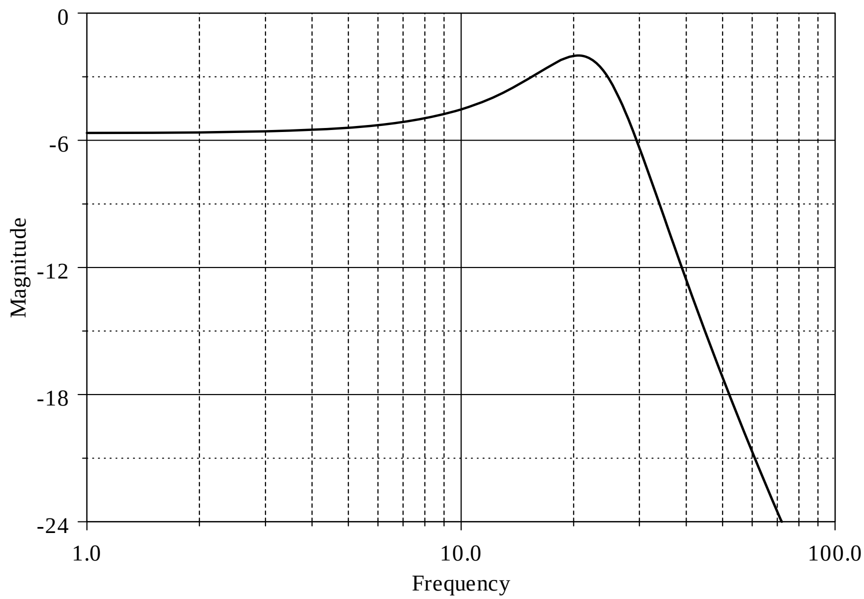


*Figure 1-7.
The dB-log spectrum of impulsive signal in Figure 1-5*

Figure 1-7 shows the same data as shown in Figure 1-6 but using the more conventional dB-log scale. Figure 1-7 is exactly the kind of graph that I will be showing often in this text—the spectrum of a signal (the signal in this example being shown in Figure 1-5). In the next section, I will describe how if this spectrum is found at the output of a system when the input was a unit **impulse**, then it is known as the **impulse response** or equivalently the frequency response. The frequency response of a system being, perhaps, the single most important characteristic of an audio component.

Impulse
a very large amplitude but very short time duration signal where the total area is one.

Impulse response
the time domain signal that would be seen at the output of a system if there was an impulse at the input.

FFT
an efficient computer algorithm for performing the Fourier Transform.

The examples that I have shown in this chapter were all generated on a computer using what is know as the **Fast Fourier Transform** or **FFT**. The FFT is a computer algorithm that aims to numerically approximate the Fourier Transform—a mathematical construct that I described earlier with an unrealistic time window of infinity. If properly implemented, the FFT will yield a close approximation of the actual Fourier Transform that we desire. Usually, any errors that can occur with the use of the FFT approximation are benign, but sometimes these errors can actually obscure what we are trying to see. In Appendix IV - Acoustic Measurements, I discuss some of these issues in more detail. For our purposes here, however, we can think of the FFT as simply being the "real world" implementation of the Fourier Transform.

At this point I would like to review the basic principles discussed in this section because they are key to the reader's understanding of almost everything that I will do in this text. These principles are described below.

- *A time domain signal has a completely equivalent representation in the frequency domain.*
- *The rules for the moving between the two descriptions in the time and frequency domains come directly from the requirements of the Fourier Series and the Fourier Transform.*
- *The FFT is a computer algorithm that attempts to implement the Fourier Transform numerically. It creates a good approximation of the true spectrum, but not without some error.*

1.2 Systems

System
a fundamental assembly block which processes input signals to create output signals.

Time invariant
a system whose properties do not vary in time.

Linear
a system whose input/output transfer characteristic is a straight line.

Now that we know how to decompose a complex signal (one that is composed of multiple tones) into a composite of simpler single frequency waves and how to scale these signals in a meaningful way, we can move on to a discussion of systems. A **system** is a symbolic block which receives data as an **input signal** and acts on that data to produce an **output signal**. It is important to limit our discussion to several restricted sub-classifications of systems. The first is that the system under consideration is **time invariant**— that is the system does not change over time, and the second assumption is that the system is **linear**.

The time invariant requirement is self-evident and simply means that the properties of the system do not change with time, or at least not significantly during the time frame that we are looking at them. If the system were not time invariant, then any statement made about it would be invalid only moments later. Linear means that when a complex signal passes through a system, each component in its spectrum is acted on individually, i.e. that the individual components do not interact with one another. This non-interacting characteristic is also called superposition, the principal that two signals can be superimposed on one another without affecting each other. For instance, when the 100 Hz component of an input signal passes through a linear system, its output level depends only on the input level of the 100 Hz component. It is not affected by any other components of the input signal, i.e. 200 Hz, 101 Hz, etc. This is an extremely important restriction, for without linearity, virtually all of the system theories that I use in this text fail to be valid or applicable.

Dynamic range
the range of signal levels over which the system can operate effectively.

The problem is that no system can be linear for any arbitrary signal level at its input or output. A system is said to have a **dynamic range**—i.e. the limits on the range of signal levels over which it is linear. The lower limit of this range is almost always the noise floor, which in some systems (like digital ones) is often a linearity issue. On the other end, at some signal level the system will saturate and cease to be able to correctly output signals or accept larger input signals. Good audio systems have a dynamic range of 90 dB or more (or a ratio of highest level to lowest level of about 30,000). Achieving this in electronics is relatively straightforward, but, as we will see, for acoustic signals in real rooms this is a major challenge. Despite the inevitability of all systems becoming **nonlinear** (not linear), it is convenient to assume linearity and move on from there. Later, I will step back and take a deeper look at the implications of the nonlinearity in typical audio system components. Linearity also comes up in the context of video signals, but it does not play as central a role in video as it does in audio.

Convolution
a mathematical technique for calculating an output signal from an input signal in the time domain.

What one usually wants to know about a system is how it acts on an input signal. This can be done in the time domain, but believe me, if math intimidates you, you wouldn't want to do it this way. The process involves a mathematical technique known as **convolution**, which is all I am going to say about it. Fortunately analyzing a system in the frequency domain, especially in dB, is relatively straightforward.

If we assume that a system is linear, then individual sine wave components in the input spectrum can be treated independent of each other. This allows a system's response to a complex waveform to be determined by simply evaluating the response of each of the individual input components exclusive of any of the other components. The components are then recombined at the output to form the output signal. This feature of a linear system is the principal motivation for moving into the frequency domain since no such simplification occurs in the time domain.

Gain
the ratio of the signal level seen at the output and the input.

Transfer function
the gain and phase of a system versus frequency.

White noise
a random signal with a flat spectrum.

Pink Noise
a random signal with a spectrum that falls at 3dB per octave.

Each input component is affected by the **gain** and phase of the system at that frequency. The gain is the ratio of the magnitude of the output signal to the input signal and the phase is the difference in the phase of the signal at the output relative to the phase at the input. In dB terms the gain is the difference, in dB between the input and the output (because division of two numbers is the same as the difference in their log values see Appendix II - Logarithms)

A plot of the gain and phase factor for all frequencies (of interest) is called a **transfer function**, $T(f)$ which is also called the frequency response of the system. The transfer function is defined as the ratio of a systems output spectrum $B(f)$ to its input spectrum $A(f)$ as

$$T(f) = \frac{B(f)}{A(f)}$$

If, in this equation the input signal were to contain all frequencies (of interest) at a unit amplitude, i.e. $A(f)=1$ then $T(f)=B(f)$—i.e. the transfer function $T(f)$ would simply be the spectrum seen at the output, $B(f)$. For example, if Figure 1-7 was the spectrum seen at the output of a system when the system had a flat spectrum $A(f)=1$ at its input, then this curve would be a plot of the systems transfer function—its frequency response. It is a classic example of a low pass filter with unity gain at low frequencies.

From the above example we can see that it would be useful to have at our disposal signals which simultaneously contain all frequencies, because placing these signals at the input to a system yields a spectrum at its output that is its frequency response. There are actually several signals with this feature. I have shown two of them in Figure 1-8. It has always amazed me that white noise (random noise with a flat spectrum) and a unit impulse have exactly the same flat magnitude spectrum. The only difference in these two signals is the phase, which is random for the white noise and a straight line for the impulse. This occurrence is a classic example of why we should always consider the phase, because looking only at the amplitude spectrum these two dramatically different signals would be indistinguishable. The term **white noise** comes from white light, which contains the full spectrum of colors, equally represented. White noise contains the full spectrum of sound of uniform amplitude. **Pink noise** has a slight emphasis on the low frequencies—the color red—hence the color pink.

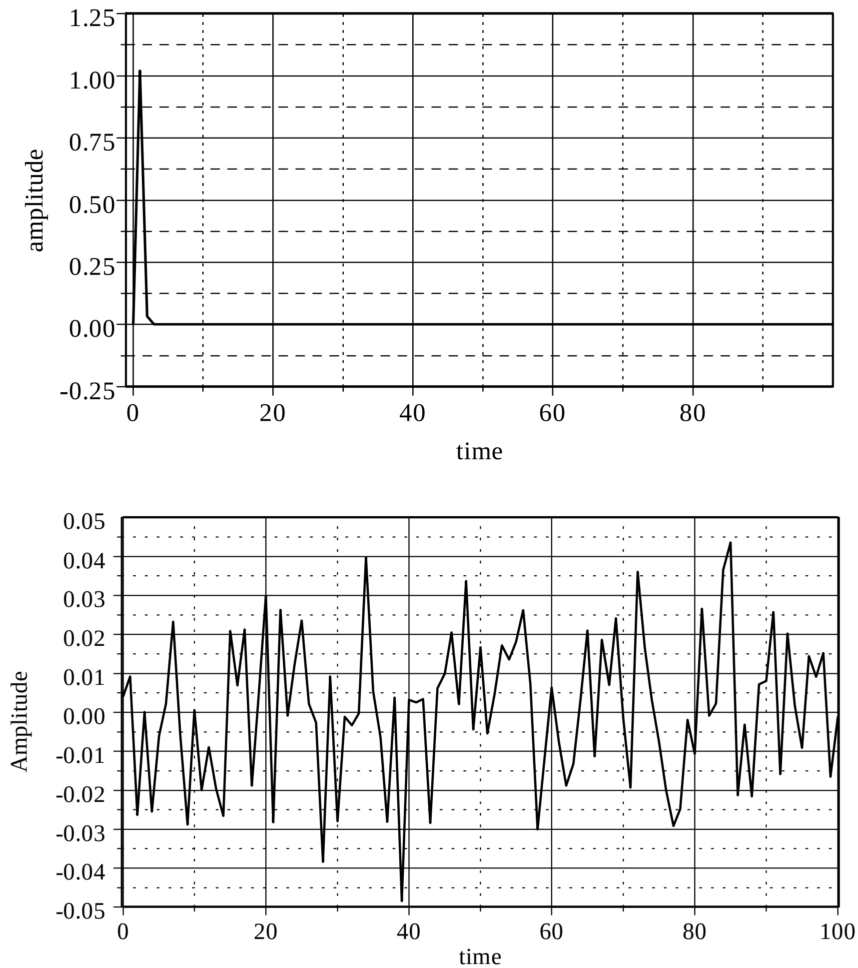


Figure 1-8. Two different signals with identical amplitude spectrums.

A third signal with a nearly flat spectrum is a swept sine wave. The details of using swept sine waves is too complex to get into here, but they are one of the most common and useful of all of the measurement signals and techniques.

Figure 1-8 shows an important characteristic of noise versus an impulse: the instantaneous level of the impulse is very high since all of the energy is located at a single point in time while the noise spreads the energy uni-

Power spectrum the frequency distribution of the power—no phase consideration.

formly across time. The impulse has a very large level for a very short period of time while the level of the noise is low for a long period of time. Both signals have the same total energy and **power spectrum** (the spectrum of the signals independent of its phase), but the peak value of the impulse is about fifty times greater than the noise. If I send either of these two signals through a linear time invariant system, the output spectrum will be the frequency response of that system. If the system is nonlinear, then it will react quite differently to these two signals which is why the linearity assumption is so important.

The response of a system to an impulse signal is logically called its impulse response. The impulse response, in the frequency domain, is the system response to all frequencies, which is its transfer function. Thus, the impulse response completely defines the linear operation of the system. If we know the impulse response of a system then we can calculate its effect on any signal that we feed this system. Interestingly, using the response to the white noise signal can also give us the impulse response, although not directly. Since the noise is random we need to do a statistical correlation of the output to the input and to get a good estimate of the spectrum, we need to take many averages. Characteristics which are not true of the impulse. The impulse, however, is likely to overload the system if not handled properly. As I mentioned, there is also the simple swept sine wave technique which can also give us the frequency response of a system. In practice, each of these methods has pros and cons as actual test protocols. The thing to remember is that if one wants to describe a system, any system, then the most important characteristic to consider is its impulse response—regardless of what signal is actually used to obtain this response. And, if done correctly, they should all yield exactly the same answer.

1.2.a Filters

Filters are one of the more common types of system manipulations that occur. This subject in its entirety is enormous and I will only touch on some of the basics.

Generally, systems do not pass all frequencies equally, whether desirable or not. Often they have a low frequency limit of usefulness (they don't always pass DC signals) and always have some high frequency limit. Systems can also be made to intentionally pass specific sets of frequencies and

Filter
a system which passes some signal frequencies while blocking others.

to reject others. Systems with these characteristics are called **filters**. Filters come in many shapes and sizes which can take many descriptors (parameters) to describe.

The first characteristic of importance to a filter is its type. A filter that passes only high frequencies is called a **high pass** filter, one that passes low frequencies a **low pass** filter and one that blocks both high and low frequencies is called a **bandpass** filter. There is also a **band reject** filter which passes both highs and lows but cuts out certain frequencies in the middle. It should be apparent that only high pass and low pass filters are unique—bandpass filters are made up of cascaded high and low pass filters while band reject filters are made up of parallel high and low pass filters. In a bandpass filter the low pass filter is set higher than the high pass filter and in a band reject, the opposite is true.

Order
the order of a filter denotes the steepness of its slopes.

Cutoff frequency
the frequency where a filters response has dropped to half power, or –3 dB.

The next most important characteristic of a filter is its **order**. The simplest filter—first order—can be described simply by its **cutoff frequency**, the frequency where the response is down by 3 dB, or one half power. Higher order filters have sharper slopes in their cutoff region. The order of a filter is determined by the number of simple filter stages in its makeup. Each simple filter stage achieves a 6 dB/octave slope which can be either positive or negative depending on the filter type. Equivalently, it has either an f or $1/f$ slope in the linear (non-dB) domain. This can be understood by recalling that the slope of a line in a log-log plot (see Appendix II - Logarithms) is the same as the power of its variable. Since sequential filters (systems) add their responses in the dB domain (multiplication in the linear domain) the filter slope increases by 6 dB/octave (or a power of f) for each filter section. Thus a three section cascaded low pass filter (order 3) would have a cutoff slope of –18 dB/octave (or $1/f^3$).

Filter Q
the shape of the filter near its cutoff frequency.

Two stage (order two) filters are described by a well-known set of descriptors; cutoff and **Q,** both of which are defined in the frequency domain. Cutoff specifies the location of the filter's slope change form its passband to its stopband and Q defines the shape of this transition. Higher order filters are often referred to by names, like **Bessel** or **Butterworth** or **Linkwitz-Riley** since these higher order filters would otherwise take too many parameters to describe. I would like to point out, however, that "named" filters are nothing more than specific cases of the general filter form, which is actually much more flexible. These named filters have usu-

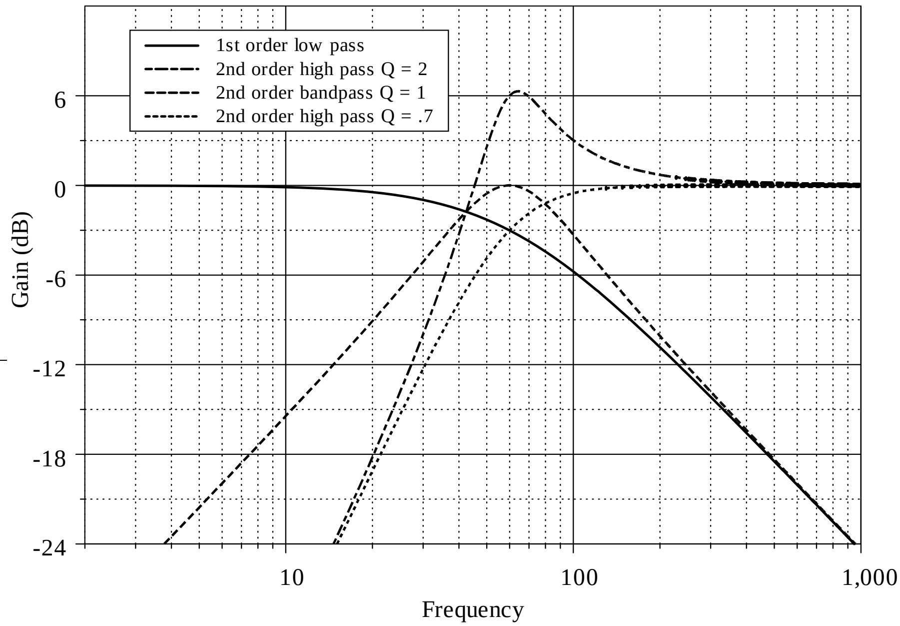


Figure 1-9. Filter function with various orders and shapes.

ally been developed with a particular goal in mind when their unique set of parameters is determined. They may have the minimum delay or be maximally flat, but otherwise there is nothing magical about them.

Some typical examples of filters are shown in Figure 1-9. These filters all have the same mathematical cutoff frequency (60 Hz) which shows that the higher order high *Q* filters do not actually have the –3 dB point at their theoretical cutoff, although if *Q* = .7 then the cutoff will be at –3 dB. Bandpass filters are defined by the sum of the high and low pass orders and are centered on the cutoff frequency with a bandwidth defined by their two –3 dB points (about 60 Hz in the above example).

1.2.b Amplifiers

Probably the most common system that we encounter after the filter is the simple amplifier. There is an enormous mystique about amplifiers, types (tube or solid state), class, etc. I think that there is a lot of evidence that amplifiers can sound different, but there is also a lot of evidence to say that any good amplifier is certainly adequate.

As a system there are usually four parameters to consider, which place the limits on its capabilities two different dimensions. The first dimension is frequency and the limits are its upper and lower –3 dB response points. As I said, no system can have an infinite bandwidth and so one needs to know what these frequency limits are. For most high quality audio amps these values are virtually always sufficient, say 20 Hz – 20 kHz, or more.

The other two parameters frame the upper and lower signal levels that can be utilized in this system component—the dynamic range. The upper level is the most commonly referenced one and is usually expressed in Watts, which unfortunately, is not a very useful number. What we really want to know is at what voltage the amplifier will clip the output signal or sound objectionable. The clipping voltage is usually fixed, while the Wattage capability depends heavily on the speaker's impedance, which is complex—literally. Always remember that clipping is a significant sound quality problem. One should never allow an amp to clip in practice for any signal as this will usually result in poor sound quality.

The final parameter is, I believe, the most important and that is the noise floor. One has to be careful here since by noise floor I mean two different phenomena. The first is the actual random (thermal) noise that appears at the output which is usually specified as a ratio of the maximum signal level to the noise level as a dB number. But, there is a second criteria, which is far more difficult to determine, and that is how well the amp handles very small signals. These small signals may be above the noise floor, but still small enough that they are affected by zero crossing errors. This type of problem can usually be seen as a rising distortion value for small signal levels. Reject any amplifier with a rising distortion for small signals as these types of distortion can be shown to be highly objectionable.

Simply put, the amplifier is, in my opinion, not the place to spend an inordinate amount of money. Any well designed amplifier will relegate this piece of equipment to the status of "not the weakest link". For example, cutting the expenditure on speakers to buy expensive amplifiers is just plain foolish. You'll only end up with a lot more power to drive your speakers into even more distortion. If you have lots of money and love the looks of a cool amp (and let's face it, this is one of their biggest appeals) then go right ahead. But please don't waste money on an amp if that forces you to cut costs somewhere else.

1.3 Video Signals and Systems

Video signals and the systems that process these signals are completely different than audio signal and yet they are amazingly similar. The first obvious difference is that a video image is a two dimension signal. However, the electronics that process these signals are basically one dimensional systems like those that I have been talking about. It is possible to do two dimensional signal processing directly using lenses or complex electronic array systems, but these are not so easy to understand and their usage is not very common.

Spatial frequency
a 2D concept where the light variations in space are analyzed as a frequency distribution.

The key concept that one must understand is that of **spatial frequency**. Once the reader has a grasp of this concept they will find that they already have the necessary background to understand image processing.Figure 1-10 shows a spatial frequency in the *x* direction of three. Compare this figure

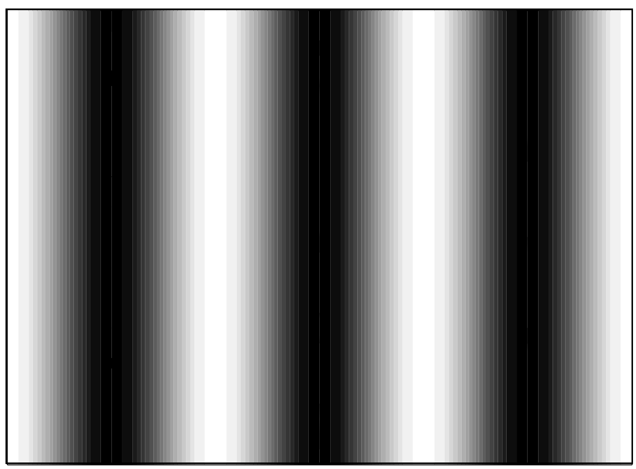


Figure 1-10. A spatial frequency of 3 in the x direction, there is no y variation.

with Figure 1-1 for the time domain. While the two figures do look different (the first figure has three waveforms), they are basically the same thing along any horizontal line in Figure 1-10. Note that there are exactly three repeats of the color (black) bars. In a video signal, on a computer for example, the values in this figure would range from 0 to 255 (for a 8 bit number in gray scale).

Wavenumber
the number of wavelengths in a given length.

Figure 1-11 shows a spatial frequency of four in the *y* direction. It is common in optics to use the **wavenumber** *k* and denote the *x* and *y* compo-

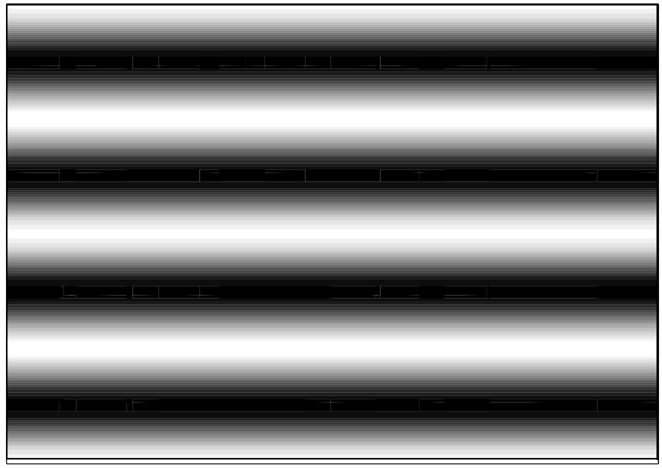


Figure 1-11.
A spatial frequency of 4 in the Y direction.

nents by k_x and k_y. The wavenumber is the number of wavelengths in a given distance. If the distance is defined to be the image width and the image height then the wavenumber are the same as the frequency.

A combination of the two wavenumbers is also possible and one such image is shown in Figure 1-12. Note that along the *x*-axis there is only a single wave and that along the *y*-axis there are two. Along the diagonal there are three

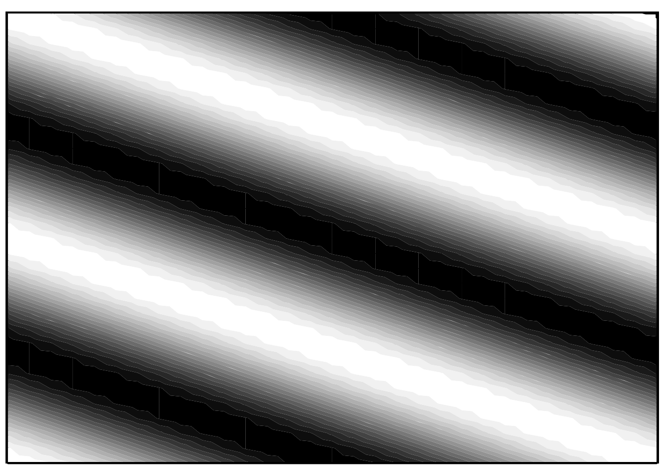


Figure 1-12.
An image with a spatial frequency $k_x = 1$ and $k_y = 2$. The total spatial frequency is then $k = k_x + k_y = 3$

Once one has grasped the concept of spatial frequency, it should come as no surprise that the most important tool for dealing with images is the two-dimensional Fourier Transform. This transform takes any image into the spatial frequency domain and in fact can be operated on by the same sorts of filters that we have already been talking about. This is because the two-dimension FFT is nothing more than a bunch of one-dimensional ones all strung together as a single steam of data. To be more accurate for image processing one usually uses the a cosine transform which is nothing more than a simplification of the Fourier Transform which makes the processing a bit faster.

I will return to this topic in Chapter 7 when I talk more about image processing.

1.3.a Conclusion

This has been a long and challenging chapter, but the principles that I have discussed here will be used time and time again throughout this text. If on first reading the material is unclear, then perhaps a second reading is in order. I am always amazed when people think that if they don’t understand something the first time they read it then they won’t ever get it. From my experience this is not at all the case. There have been many texts that I have had to read twice to understand and I know of certain sections in texts that I have read four or five times to truly comprehend. There is no doubt that technical subjects can be difficult to follow, but diligence always pays big dividends in the end.

PART I

AUDIO

CHAPTER 2 HEARING AND PSYCHOACOUSTICS

WITH LIDIA LEE

I would like to lead off the specific audio discussions with a description of the audio receptor—the ear. I believe it is always a good idea to understand some basics of what these receptors are capable of, but equally or more importantly, what they are not capable of, before discussing design details of systems that utilize these receptors as their target.

2.1 The Ear

In this chapter, the key concepts in psychoacoustics that are critical to later chapters will be discussed. Understanding of psychoacoustics requires a basic understanding of the theory of hearing. Hearing is an incredibly complex multidimensional system which is composed of acoustical, mechanical, hydrodynamic and neurological subsystems all acting basically in series. Each one of these subsystems affects the perception of sound, but some more than others. Figure 2-1 shows a schematic drawing of the ear where three prominent subsystems called the **outer ear**, the **middle ear** and the **inner ear** are seen.

Outer, Middle, Inner Ear the three fundamental subdivisions of the ear.

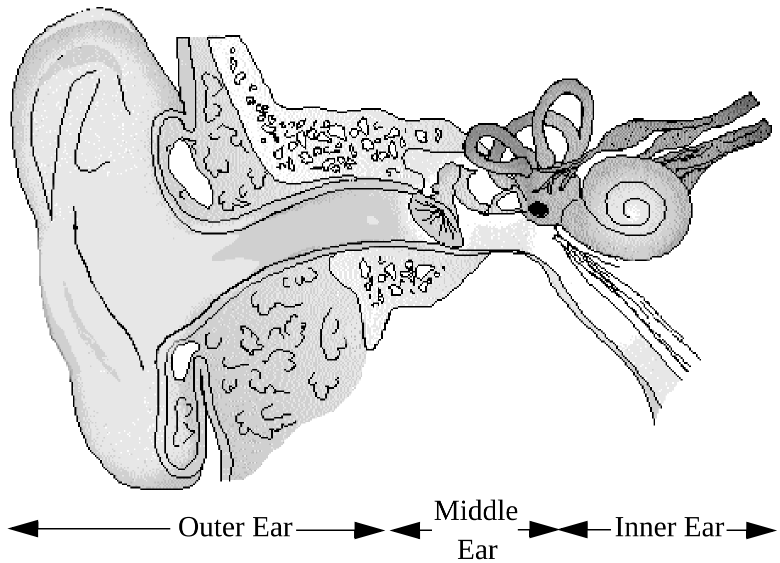


Figure 2-1. Schematic drawing of the ear system with three subsystems

2.1.a The Outer Ear

Binaural
the use of two ears.

ITD
the time difference between sound arrivals at the ears.

IID
the intensity difference between the sound at the ears.

The effect of the **outer ear** is primarily an acoustical one. Combined with the head, the outer ear forms a spatial acoustic filter which acts on sounds coming from different directions in different ways. The **binaural** hearing system (mostly responsible for sound localization) is composed of two ears, the signals coming from these two receivers and the signal processing that takes place in the brain. Sound localization is a function of two basic characteristics of the ears. The first is the time delay from one ear to the other which is called the **interaural time difference (ITD)**. There is also a level difference between the ears due to the head shadow effect which results in an **interaural intensity difference (IID)**. The brain acts on these two signal parameters to determine the location of a sound source, and it does so in slightly different ways depending on the frequency.

The Duplex Theory of localization proposes that low frequency signals are detected by ITD, and high frequency signals are detected by IID. The distinction of low versus high frequency is determined by the dimension of the head where the dividing frequency is approximately 1 kHz for a head diameter of approximately 8.5" —the range above is considered as high frequency and the range below is considered as low frequency.

IPD
the phase difference between the sound arrival at the ears.

The ITD together with the **interaural phase difference (IPD)** provides pertinent low frequency localization information. Without the critical IPD information, temporal confusion would occur and hinder the ability to localize low frequency sound sources. Assume that a signal approaches from the left side (270°) in the horizontal plane. The stimulus will arrive at the left ear before it reaches the right ear, hence creating a time difference between the two ears. As the signal moves from the left (270°) to the front (0°), or to the back (180°), the ITD is indistinguishable, that is, the signal arrives at both ears at the same time if presented at either location. Hence, it is difficult for us to localize sound sources at many locations based solely on ITD. In a reverberant room, studies have demonstrated that a person's ears can systematically identify the first wavefront that arrives at the ears and localize the sound source—rejecting the confusion caused by the many reflected secondary signals. This phenomenon is known as the precedence effect or the law of the first wavefront. It should be noted, however, that nothing is implied about tone color changes or increased source location confusion that may occur with these reflections, only that the sources location is principally controlled by the first arrival. This later aspect of the precedence effect is often misunderstood.

The IID refers to the difference in signal level between the two ears. It is created by the diffraction effects of the human head at mid frequencies and the pinna effects at high frequencies. Sound localization is dependent on the ITD at low frequencies and the IID at high frequencies. At low frequencies, there are only small differences between the signals at the two ears, no matter where the sound source is located, which results in small ITD value and poor localization at low frequencies.

Since the two ears lie in the horizontal plane, the interaural time and intensity differences for sound sources in this plane are at a maximum. For sources in the vertical plane the ITD and IID virtually vanish. This means that the sense of localization is greatest in the horizontal plane (just as we might have expected of nature since the world lies mostly in this plane).

Thus, it can be seen that the acoustic filters of the outer ear accounts for much of one's localization capabilities, but the brain does do the final determination.

2.1.b The Middle Ear

Cochlea
a spiral organ lined with receptors, fluid filed, which is the main detector of sound.

The next subsystem in the chain is primarily mechanical consisting of the ear drum and small bones connected to another small membrane that is attached to the fluid-filled **cochlea**. The principle function of the middle ear is to transform the small acoustic pressure vibrations in the ear canal into a fluid motion in the cochlea. The bone structure of the middle ear can also act as a slow reaction time sound compressor that results from changes in the flexibility of the system with sound level. The three bones in the middle ear (the smallest in the body) are held in place by muscles and constitute a mechanical lever. At high sound levels these muscles tighten, thereby stiffening the lever. This effect is used as a protection device against sustained loud sounds, and results in the occurrence of the temporary threshold shift that we have all experienced after being subjected to loud sounds for a sustained period. Unfortunately, this protective mechanism cannot react fast enough to protect from sharp impulsive sounds like gun shots, etc. This is why these sounds are the most damaging to the hearing system. No doubt, the lack of naturally occurring impulsive sounds is the reason for this lack of protection. Virtually all impulsive sounds in the environment are man made (thunder being one notable exception), and, unfortunately, they are abundant in the current environment.

2.1.c The Inner Ear

The inner ear, which is principally hydrodynamic in its action, is by far the most complex system that will be discussed in this text. Even today its function is not completely understood. (Just imagine how difficult testing this system must be! It is extremely difficult to test directly on live subjects and its function deteriorates within minutes of death.)

Basilar membrane
a membrane in the cochlea which contains the principle hair cells for sound detection.

The inner ear starts with the **oval window** at one end of the cochlea which is excited by the middle ear’s mechanical vibrations. This excitation creates a wave that propagates into the fluid-filed cochlea. The cochlea consists of a long tapered tube which is divided in half by a membrane and coiled up in a helix. The dividing membrane is called the **basilar membrane** and its motion triggers small hair cells embedded in it which activate nerve impulses that travel to the brain. The exact action of this motion and the nature of how the neurons fire is far too complex to get into in detail here;

however, an overview of the general theory of how wave motion on the cochlea is translated into the perception of sound will be discussed.

2.2 Theories of Hearing

2.2.a Place Theory

Place Theory is considered to be the most well established to date. This theory suggests that the cochlea is partitioned into different segments along its length. There are thousands of these segments lying along the length of the cochlea, which, because of its constantly changing size and shape, forms a sort of place-tuned resonant system. Each segment thus responds to a certain limited band of frequencies. As a complex signal reaches the cochlea, a form of mechanical frequency analysis takes place as the waveform travels along the cochlea. This analysis is similar to a Fourier analysis, however significant difference exist. It is thus erroneous to think of the cochlea as a Fourier Analyzer. The different components of a complex signal excite the cochlea to a maximum amplitude of vibration at different points along the cochlea depending on the frequency of these components of the signal. The cochlea senses higher frequencies at its input end and lower ones at the opposite end. Thus, the high frequency hair cells are excited by all frequencies—a reason why one tends to loose the high frequency hair cells first, which explains the domination of high frequency hearing loss in the general population.

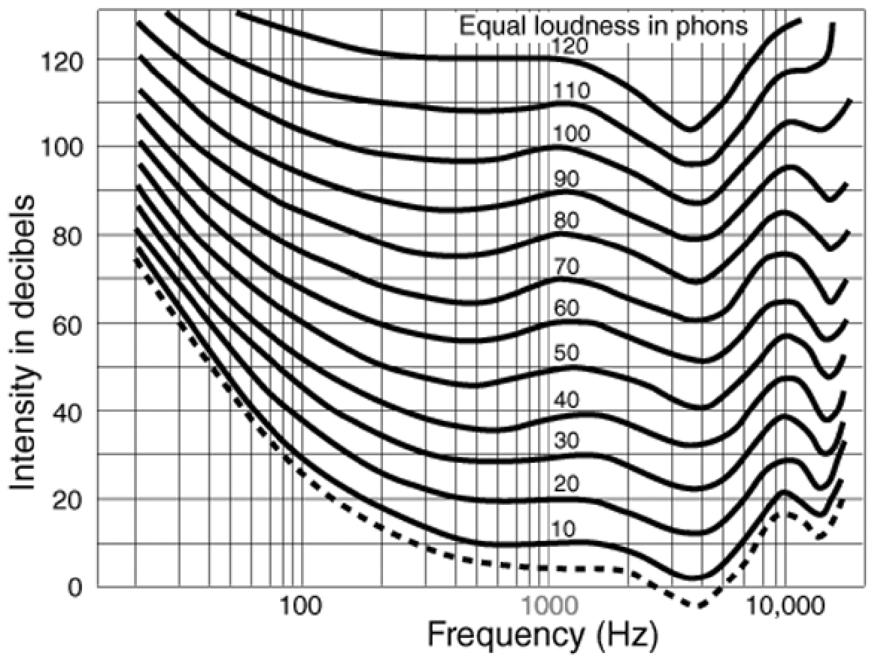
2.2.b Frequency Theory

Frequency Theory is based on the assumption that the auditory nerve fibers fire at a rate which is synchronous with the input signal, transmitting this information to the brain. For example, a 100 Hz signal is detected by nerve firings at a rate of 100 times a second. The intensity of the sound is coded by the number of impulses in each volley. This theory is adequate at explaining the perception of low frequencies, however, above about 1 kHz it becomes untenable due to a nerves finite recharge rate. The recharge rate is the time it takes an auditory nerve cell to re-establish its polarization in order to fire again and is typically about 1 ms long (which corresponds to

1 kHz). Hence, above 1 kHz, the neurons are simply unable to fire synchronously with the input signal. They cannot keep up, and Place Theory becomes the dominate perception mechanism.

A look at Figure 2-2 shows some interesting results of these two theories. This figure has several curves, each of which represents an equal perceived loudness set by their value at 1 kHz. Above 1 kHz, we see a basically flat response which Place Theory would suggest. There are of course various resonances in this response due to the acoustic resonances of the outer ear structure. The dip at about 3 kHz is caused by the ear-canal resonance and a second resonance occurs at about 14 kHz.

Figure 2-2. Contours of equal perceived loudness.



Below about 500 Hz, there is a direct result of Frequency Theory where the perception of level is strongly dependent on the frequency and the input level, but less so in the later case. This is easy to explain with Frequency Theory. Consider that; as frequency drops below about 500 Hz, there are less and less nerve firing (they are synchronous with the frequency) and since the loudness is perceived by the number of impulses, the perception decreases with frequency (a rising curve in Fig. 2-2). Further, as the level increases there are more nerve impulses per volley (because of the higher hair cell excitation). Thus, the curve flattens out at higher levels of low frequency simply because the perception of loudness is proportional to the number of nerve firings in any given period.

2.2.c Place-Frequency Theory

The prevailing theory of hearing is called **Place-Frequency theory**. This theory combines the theories of Place and Frequency into one. At low frequencies the auditory system does an actual time-domain sampling of the signal (Frequency Theory), while at higher frequencies the location of maximum amplitude on the cochlea is the primary detector (Place Theory).

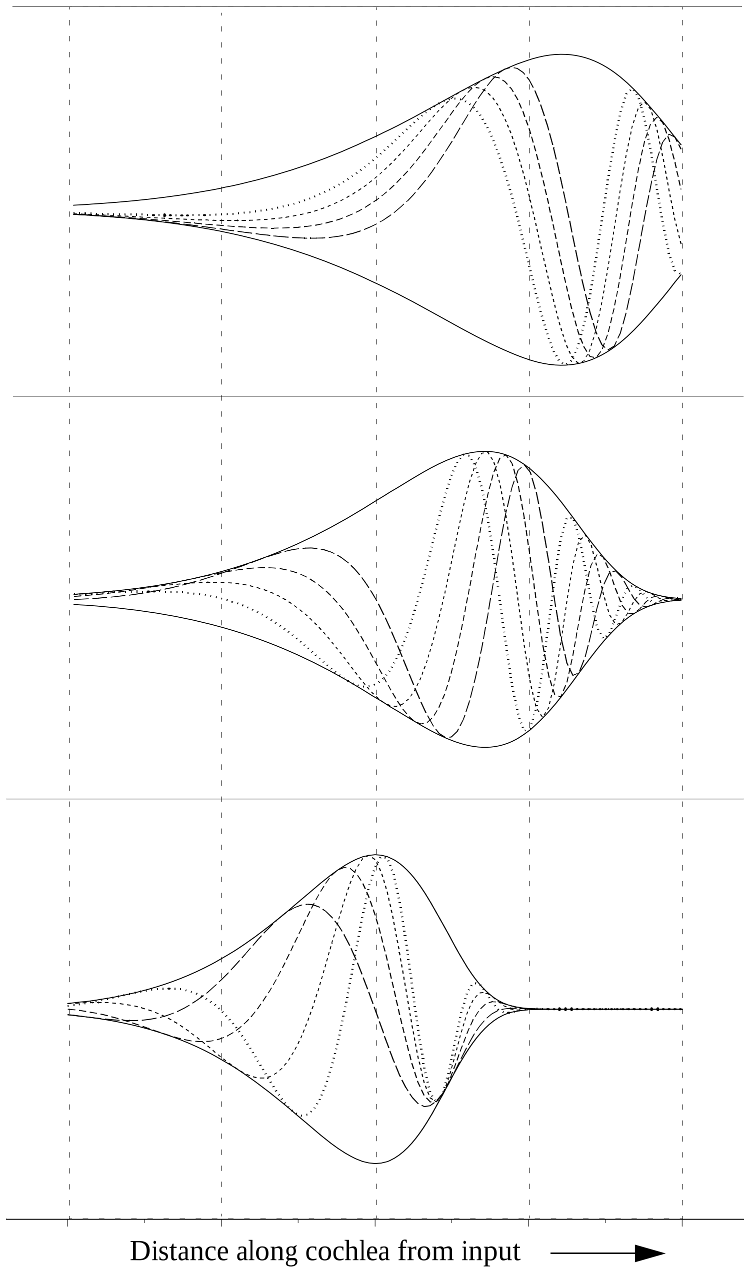
2.2.d Masking

The next concept in hearing that will be discussed is best visualized by considering the Place Theory description of wave motion on the cochlea as shown in Figure 2-3 on the next page. This figure is a graphic representation of how a wave travels along the basilar membrane inside of the cochlea. The solid line represents the envelope of the wave motion and the dotted lines are instantaneous wave amplitudes. It is readily apparent that for frequencies above a few hundred Hertz, there is a strong place or location component to the amplitude of excitation. At low frequencies, this place aspect is becoming less apparent and the frequency theory component of hearing takes over as the dominant factor in perception. Note that the frequency "place" on the cochlea moves towards the input end at high frequencies of excitation. In this figure, the frequency place is the opposite of what we are used to seeing where high frequencies are to the right.

The most important aspects of these curves are the tails in the envelope that extend both above and below the peak. The lagging tail, towards higher frequencies, always has a greater extension than the leading tail, to lower frequencies. Acoustic excitation that lies "inside" these tails is said to be masked. These frequencies are masked simply because the ear cannot perceive signals which lie within one of these tails. Masking theory is probably the single most important concept in psychoacoustics as it applies to audio. It will show in the next chapter that masking accounts for most of the enormous data reductions that are now possible with **perceptual audio coding** techniques such as MP3 or WMA (see Chapter 3), and it also has a strong

Perceptual Coding
the use of perception as a basis for data reduction techniques.

Figure 2-3. Graphic representation of cochlea traveling waves at different frequencies: 100 Hz (top), 400 Hz (middle), 800 Hz (bottom)



effect on the perception of distortion. Distortion products cannot be heard when they are masked.

Masking predominantly affects higher frequencies of the original sequence and the term “upward spread of masking” is usually used. Masking is also level dependent, which is not apparent from Figure 2-3. Figure 2-4 shows measurements of the masking curves for various signal levels of a 500 Hz tone (Denoted as the numbers on the curves 20–80 dB). A signal that lies below the pertinent masking curve in this plot would not be audible. Note that the masking spreads upward in frequency to a greater extent at higher levels and that the masking curves get wider at higher levels. This effect also impacts the perception of distortion since distortion byproducts tend to be upward in frequency and will be masked to a great extent.

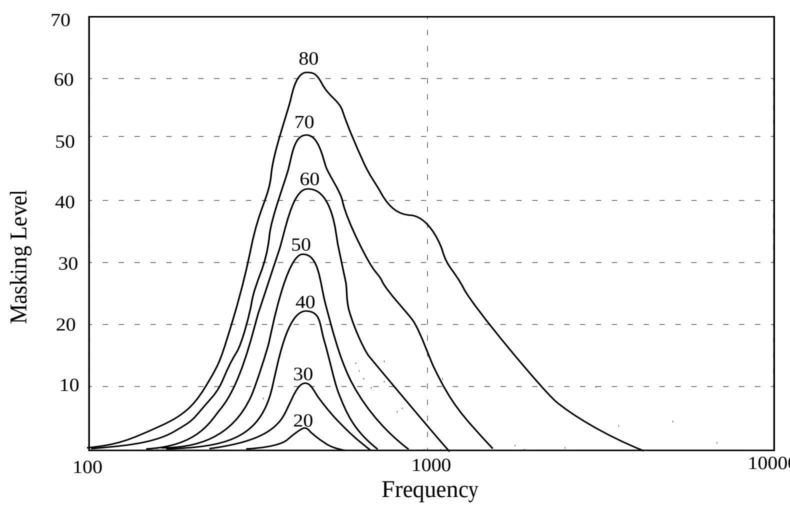


Figure 2-4.
Graphic representation of masking patterns of a fixed probe frequency. Each contour represents a different probe intensity.

For a simple single-tone excitation, it has been shown that the perceived loudness is a complicated function of the frequency and level of the excitation. More importantly, the perceived loudness of a tone in the presence of other tones can vary from normal to completely inaudible depending on the level and location of the other tones. The net result is that the perception of a complex pattern of sound excitation can be difficult to analyze and understand. It should come as no surprise that simple acoustic measurements of sound levels are a long way from giving an accurate description of how a sound system is perceived.

Since the waveform on the cochlea takes a finite amount of time to build up and decay, there will also be a temporal masking effect, but this effect is not as great as frequency masking. More important, in the temporal domain, the ear has a finite integration time within which discrete temporal events are fused together into a single excitation pattern. This later effect is of fundamental importance in small room acoustics.

2.2.e Critical Bands

Critical band
the effective bandwidth of an auditory filter.

A direct result of the finite width of the wave-packet motion on the cochlea, as shown in Figure 2-3, is that the ear has a finite capability to resolve signals in the frequency domain. The extended width of the wave packet simply does not allow for a precise resolution of signals with components that are close together in frequency. As a system, the ear acts as if it had a multiplicity of bandpass filters operating in parallel. The bandwidth of these filters is called the **Critical Band**. It must be stressed that the critical band filters themselves are not actually fixed in location, rather they are more like swept filters. It is convenient in discussions about psychoacoustics to consider the ear as having a frequency domain resolution that is equivalent to a set of about 30 bandpass filters covering the auditory spectrum.

Fused
the effect of two sounds arriving within a small time window being detected as a single sound.

The critical band is defined as the frequency bandwidth within which different frequency components of a signal integrate together—the term **fused** is sometimes used. Individual tones within a critical band are not perceived individually but as a single tone at a single level. In other words, two tones within a critical band will not have an additive loudness characteristic as two tones separated by more than a critical band will. The bandwidth of these critical band filters is frequency dependent as shown in Figure 2-5.

The critical-band concept is important because it implies that those portions of a signal that lie within a common critical band can be treated as a single composite signal. Each critical bands composite signal then interacts with the other critical bands of auditory filters as simple composite signals and not as complex signals. This concept will be shown to be necessary for the audio compression techniques described in the next chapter.

It is important to realize that even though we have a finite critical band resolution in the frequency domain, we can still resolve the frequency of a signal within a critical band to a resolution of about 5% of the bandwidth of

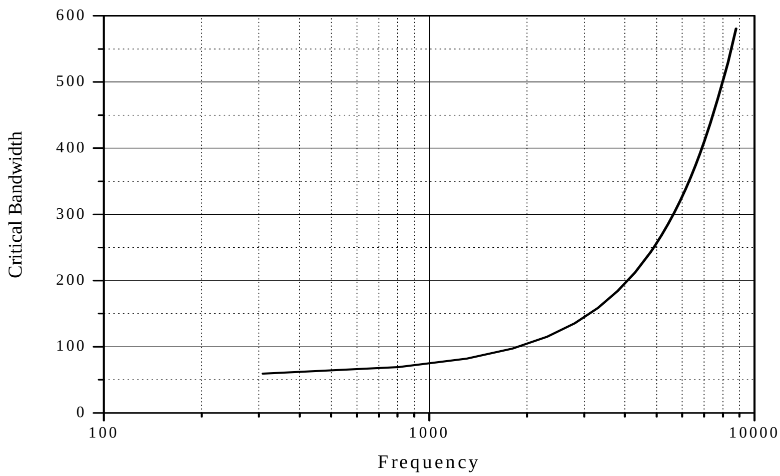


Figure 2-5. Critical bandwidth as a function of frequency.

the critical band. Thus, the critical band concept does not alter the perception of pitch and perfect pitch is quite possible. The critical band concept is only useful when talking about complex signals with a multiplicity of tones present. When two or more of these tones lie in a critical band, they must be treated differently than multiple tones that lie outside of a critical band. This is a particular instance where the Fourier analysis concept fails for the hearing mechanism.

2.3 Summary

To design optimum AV systems, one has to know what the target resolutions of human receptors are—in real terms—for it is unwise to spend time or money chasing after criteria which are not perceivable in the end product.

There are several main features of the human auditory system that you should have learned from this chapter.

- The ear has good spatial resolution in the horizontal plane at mid to high frequencies, but at low frequencies its spatial resolution is poor.
- The ear has a spatial resolution in the vertical plane that is poor because of the equal distance from each ear to points in this plane.
- The ear masks smaller signal components in the presence of larger signal components. This effect acts greater on signal components whose frequency is above the main signal component (the masker) than below it. The overall effect increases with the level of the masker.
- The ear also masks smaller signal in time making a small signal which follows a larger one in time inaudible.
- The ear has a finite frequency resolution which can be approximated as a bank of bandpass filters where the width of each filter is known as the critical band. A critical band's width varies with frequency.
- The ear fuses complex sounds that lie within a critical band into a single perceived tone.

These features will be used in the following chapters to define and quantify numerous aspects of signal compression, audio system sound quality and room acoustics.

CHAPTER 3 AUDIO MEDIA

3.1 Background

It is important to understand how the various audio software is distributed in order to plan for its use. Today, there are so many audio media formats that sorting them out is no small task. Years ago it was simple—analog on either a vinyl disk or magnetic tape. The two media were roughly equivalent although they both had different (obvious) flaws. I will discuss the analog formats briefly and then the non-compressed digital formats, ending with a detailed discussion of the new field of compressed audio, which is clearly the future.

3.1.a Analog

Analog formats basically record some mechanical or magnetic characteristic which is an equivalent copy or “analog” of the original waveform. For LPs, it was the actual mechanical displacement of the groove while for magnetic it was the strength of the magnetic field placed on a magnetically coated backing. It is interesting that today’s hard disks are in fact an evolution of the magnetic tape of the early days. Some early PC’s actually used cassette tapes for their mass storage needs (when 64k was a big program). As ridiculous as this sounds, it sure beat paper tape! (A one inch wide very long strip of paper with holes punched in it.) In principle, analog recording is the highest quality since there are no conversions of the signal required at any point. However, in practice this is never the case. For LPs, the mechanical integrity of the grove is a big issue. Dust or minor scratches degrade the "image" recorded on these surfaces as does wear from the actual playback process. One could, and this has been done, sense the LP groove optically with no resulting wear of the media, but the dust and scratch problems still remain. There is also no reason why magnetic tape could not be made higher quality given the refinements in sensors and magnetic coatings that have resulted in the last few years from the magnetic storage industry, thereby yielding a near perfect analog medium.

Clearly analog storage is not an unreasonable proposition, it is simply that **digital** is now so prolific in our everyday life that analog storage is just no longer viable as a competitive technology. Add to this situation the tremendous peripheral capabilities of digital and it is a hands-down winner. What exactly are these new digital technologies that are making digital so ubiquitous?

3.1.b Digital Signals

As a subject of discussion, digital audio is enormous. By now most people are comfortable with the digital interpretation of audio thanks in many ways to the computer. When one can actually see the digitized signal with a computer program like Cool_Edit, then the understanding of how digital sampling works follows very quickly. If you have not looked at a wave file (an audio file with a *.wav extension that is an exact sampled representation of an actual acoustic waveform, we will discuss wav files a little later) with a program like Cool_Edit, then I highly recommend that you do so. Even experienced practitioners, like myself, find it enlightening to actually see what it is that we are listening to as we listen. Nothing quite compares with a direct experience like this for understanding a technology such as digital audio.

Sample Rate
the number of sound measurements (data samples) made per second.

Bits of resolution
the number of data bits in the data sample.

In the digital world there are two factors that control the quality of the recorded signal. The first is the **sample rate**, or how often a sample of the waveform is measured for storage to the recording medium, and the second is the **data word length** or how many **bits of resolution** are used to encode the value of that sample. The current standard, that found on a CD, is a sample rate of 44.1 kHz and a word length of 16 bits. An emerging standard (one in which I am dubious about) is a 96 kHz sample rate with a 24 bit word length. Later on in this chapter, I will show some strong evidence for why this higher resolution may be a waste of bandwidth (a word which I will also define in more detail a little later). Cool_Edit can use a number of different sample rates and word lengths and the reader is encouraged to experiment with these variations. Since I will be encouraging the use of a computer as the base platform for the AV sound and image processing, it is a good idea to become somewhat proficient with some of the more common applications and techniques. At the time of this writing, Cool_Edit, for example, is free with limited capabilities and is readily available on the net. The full version is also well worth the purchase price. As a base platform for audio file manipulation it is unsurpassed. (Cool Edit is no longer available having been bought by Adobe and converted into Encore. There are many Cool Edit facsimiles, but none seem as good as Cool Edit.)

Consider now a stereo signal sampled at 44.1 kHz with a 16 bit word. This signal has to pass:

$$44{,}100\ {}^{samples}\!/_{\mathrm{sec}} \times 16\ {}^{bits}\!/_{sample} \times 2\,channels = 1.4\ {}^{megabits}\!/_{\mathrm{sec}}$$

Bit rate
the number of bits sent through a channel per second.

This value is known as the **bit rate** and it means that every seven seconds a megabyte of data has to be processed. Today, this seems like an achievable task, but even a few years ago this was a major feat. From a storage perspective, this is still a huge amount of data. A 3.5" floppy can only store about 10 seconds of audio at this bit rate. And even today a fast ethernet home network could not pass this data rate without some error or compression. Fast internet hookups, like the higher bandwidth ADSL lines, can only carry this amount of data in real time with some difficulty. Clearly, passing audio around computer systems requires a substantial reduction in the data rate or **bandwidth** as it is also known.

Bandwidth
in digital data steams—the width of the frequency band that is required to carry the data.

Frequency Modulation
the technique of changing the instantaneous frequency of a waveform as a means to encode data.

The term bandwidth originated literally from the width of the **Frequency Modulation** (FM) signal that is required to transport a given signal. Generally, about 1.5 Hz of bandwidth is required for each bit in the bit stream. That means the minimum required bandwidth of the above stereo signal is about 2–3 MHz (mega-Hertz) for transmission as a digital signal. In general, virtually all data transmission is done by frequency modulated data transmission techniques—like FM radio. In FM, the carrier frequency is the center frequency about which the frequency modulates, which is how the data is encoded. The carrier frequency can be placed anywhere, as long as it is greater than the modulation frequency bandwidth. The modulation frequency, or bandwidth, tends to be fixed by the data requirements. The minimum **carrier frequency** required to do this task is about 1.5 MHz with a modulation frequency of about 1.5 MHz. We can see that the important parameter is the data bandwidth since this determines how much data a **channel** can carry. When looked at in this way, transmission of digital data are real bandwidth hogs. After all, the physical bandwidth of the audio signal is only two times 44.1 kHz or about 100 kHz. However, when the carrier frequency is 100 MHz, then 1.5 MHz modulation is not much of a problem and when the carrier is in the Gigahertz range, as in cell phones, etc. then this bandwidth is even less of a burden. Unfortunately, computer networks are not quite this fast—at least not yet. At any rate, it should be obvious that for high quality audio to be useful on a computer, some substantial reductions in bandwidth are required.

Carrier frequency
the base frequency about which the FM signal is modulated.

Channel
in digital data streams, the path taken by the data stream.

The first obvious way of reducing bandwidth is to reduce the sample rate and/or the word length. These reductions have obvious detrimental effects on the sound quality but do represent an often used technique for bandwidth reduction at the extreme. Some webcast radio stations have extremely low sample rates and word length (along with the techniques that I will discuss below) to allow them to be transmitted over the internet. For voice signals these reductions are generally benign, but for music, these techniques certainly lack appeal.

The next approach we could take is to consider compressing the actual bit stream by looking for patterns or long strings of 1's and 0's, as would often occur in an audio signal file. This type of compression is called **loss-less** because the original digital bit stream can be reconstructed exactly with no loss in data. There are numerous techniques used in loss-less **coding** (the term given to compression techniques when applied to AV signals), but in essence they work exactly like common software programs that create *.zip files (WinZip, etc.). In fact, loss-less coding can be easily performed by simply taking a *.wav file and creating a zip file out of it. Compressions vary dramatically depending on the *.wav file but, typically, only about a 10% reduction in file size is achieved with this technique. Windows Media Player has a loss-less encoding option which seems quite effective. Our media storage is all encoded in this loss-less format.

Loss-less Encoding
data encoding which returns an exact duplicate of the original

Coding
the technique of data reduction using a computer algorithm.

The next class of data compression is called **perceptual audio coding**. With these techniques, and there are currently a proliferation of them, any data that cannot be “perceived” by the listener, and hence is superfluous, is removed. Perceptual coding forms the basis for virtually all modern audio **codecs** (the term for an audio coder/decoder) in use today, so I will take some time to explain the background for these techniques.

Codec
the input and output algorithms for encoding and decoding a data stream.

3.2 Perceptual Coding

Now that the reader has a basic understanding of hearing, spectral masking (Chapter 2) and the concept of the frequency domain (Chapter 1), I will put this information to use.

The masking of one signal by another is the fundamental characteristic of human hearing that makes much of the data in an audio signal imperceptible and, hence, superfluous. Consider a signal composed of two sine waves, one at 500 Hz with a level of say 80 dB (SPL) and another at 1 kHz at 40 dB (SPL). In this example the second tone would be completely inaudible (see Figure 2-4 on pg. 37). By actually eliminating the second tone, we can reduce the data requirements by only coding the 500 Hz tone. This data analysis must be done in the frequency domain since in the time domain the frequency masking effect is not evident. There are time domain masking effects, as I discussed in Chapter 2, and these are also used in many codecs to further reduce the data, but these effects are secondary to frequency domain masking.

What about a complex signal? How can I determine what data can be eliminated and what cannot? First, consider the following situation. For a low level signal there is very little masking. Low level signals use only a few low order bits with the upper bits being set to zero. Low level signals will benefit from the loss-less coding portion of the codec. For high level passages there is a significant amount of masking, so these passages benefit most from perceptual coding techniques. When combined, these techniques lead to a fairly consistent data reduction level for the entire signal.

In an actual codec, the incoming audio signal is filtered into a number of bins using a bank of filters which are set to represent the ears’ critical bands

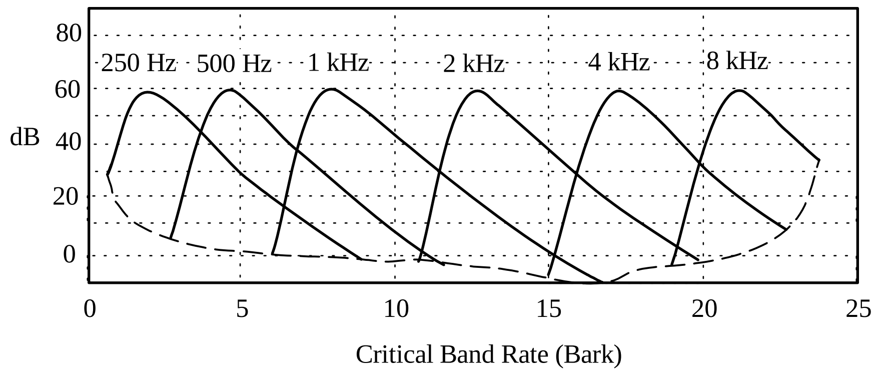


Figure 3-1. Masking effect versus critical bands.

(see Chapter 2 pg. 38). The reason for this choice of filters is that the masking effect is independent of frequency when plotted versus critical band rate as shown in Figure 3-1. The scale of this figure is in **Barks**, which is a sort of frequency warping that makes the critical band filters appear to have similar shapes. This allows for a single masking algorithm for all of the bins. There are typically 32 bandpass filters or bins set to cover the entire audible frequency range in a perceptual coder.

Bark
the unit of proportional frequency used in psychoacoustic data.

If, in our above example, I had set the secondary tone to 60 dB instead of 40 dB, then this secondary tone would now be audible; however it would only have a dynamic range of 20 dB (the level above the masking level) due to the presence of the masking tone at 500 Hz. This dynamic range, which is essentially a signal to noise ratio, is called the **signal to masker ratio**. Returning now to a complex signal passing through the codec I would calculate the masking effect from each adjacent bin to determined a signal to masker level for each of the 32 bins.

Signal to masker ratio
the ratio of the signal level to the masker level in a critical band.

Knowing the signal to masker level in each bin allows me to determine the minimum number of bits that are required for each **coefficient** (the value of the FFT at a particular frequency) in order to represent the signal at that frequency with the required signal to masker ratio. That is, the value of the frequency components in each bin are represented by numbers which have no more dynamic range than the ear is able to detect. It is well known that each bit in a coefficient contributes 6 dB of dynamic range. Rather than using 16 bits to represent the value at each frequency in a particular critical band, only the number of bits required to just exceed the signal to masker ratio are used. The reduced number of bits in a coefficient in a bin will

Coefficient
the level of an FFT bin at a particular frequency.

Quantization error the noise created by the finite bit depth in the data word.

cause data noise, **quantization error**, to appear in that bin, but this noise will not be perceived since it is masked.

In our two tone example, I would need

$$\frac{80\,dB\,(500\,Hz.)}{6\,dB} = 14\,bits$$

for each coefficient in the critical band containing the 500 Hz tone, plus

$$\frac{20\,dB\,(1000\,Hz.)}{6\,dB} = 4\,bits$$

in the critical band containing the 1 kHz tone and no bits in any of the other bins. If the FFT of the signal has 1024 points and there are 32 critical bands covering this signal, then each band has 32 coefficients. (This is not exactly correct since the critical bands are not all the same number of FFT points, but you get the picture.) The total data that would be required in this example is:

$$32\,coefficients \times 14\,bits + 32\,coefficients \times 4\,bits = 576\,bits$$

for the compressed signal versus

$$1024\,coefficients \times 16\,bits = 16,384\,bits$$

for an uncompressed signal. This is a remarkable reduction in data for little or no compromise in perceived sound quality.

This example is extreme since most of the frequency domain is empty and real signals would not usually have this advantage. Still, it is quite apparent that this is a dramatic reduction in the data required to represent the compressed signal with the same perceived signal to noise ratio as the original.

By taking an inverse FFT at the transmission reception end, I can reconstruct the data in time blocks and link them together in a continuous chain to recreate a reasonable facsimile of the original signal, from a perceptual point of view. However, It is highly unlikely that the reconstructed time signal would still look like the original.

In effect, we have a technique whereby the masking effect of the human hearing system can be used to reduce the information bandwidth required to model a perceptually equivalent signal with much less data. It is only perceptually equivalent because in the time domain the recreated signal will look very different than the original signal. The signal has in fact been significantly distorted, but in a way that is imperceptible. This brings us back to a point that I made in an earlier chapter. Altering the time domain waveform of a signal visually (and mathematically) distorts the signal, but, because of masking effects, we cannot conclude from this that this alteration causes a perceived degradation in the sound quality. Standard linear system tests for distortion, like Total Harmonic Distortion (THD) or Intermodulation Distortion (IMD), when performed on a perceptual codec can produce extremely high numbers, but I have shown how these distortions could be inaudible. The conclusion that can be drawn from this simple fact is that standard linear systems theory should not be applied to a system where perception is the basis of evaluation. If one wants to evaluate the relationship between the measurement domain and the perceptual one, they have to consider the characteristics of the hearing mechanism, the human ear. This evaluation system is not a linear system and one should not use simple linear system theory in this kind of evaluation. Perception is a highly nonlinear function.

In Chapter 1, I talked about the distortion in a system, more specifically, in an amplifier. It is important to realize that it is not the levels of THD or IMD that indicate a problem with the system, but how these systems act on the actual signals passed through them. This is a complex issue. It turns out that it is the higher harmonics of a distorted signal that pose the greatest audible deficiency and that these defects are more audible at lower signal levels. This is exactly why I stated that the distortion must not rise with lower signal levels in a system-this is a clear indication of poor sound quality.

3.2.a MP3

With the basics under our belt lets take at look at the most common perceptual coder the Motion Pictures Experts Group (MPEG) Layer 3 codec. Dating back to the early 90's the MPEG began to study ways of reducing the transmission data for movies, television, etc. We will see in Chapter 8 how this group has also defined the current standards for video data reduction. In this chapter, I am primarily concerned how they defined the standards for audio reduction.

MPEG defined three alternatives for audio compression, called layers:

Layer 1) uses a single data block of 384 samples in 32 critical bands where only frequency masking is used.

Layer 2) uses three data blocks, total of 1152 samples, which does a cursory time domain masking simulation.

Layer 3) uses an improved auditory filter model (critical bands) which vary with frequency. It also adds in a more sophisticated time domain masking model and does stereo redundancy reduction. The data is finally reduced with a bit reduction scheme.

Layer 1 has about a 4:1 reduction, layer 2 about 6:1 and Layer 3 an impressive 12:1. It should be noted that MP3 is one of the earliest codecs and that many newer algorithms are known to work better. Still MP3 remains as one of the most prevalent and its use is extremely widespread.

3.2.b Windows Media

Microsoft has one of the leading algorithms from a perception standpoint—WMA. It is currently part of the Windows Media Player and has several fixed bit rates, a **variable bit rate** and a loss-less option. One can also create.wav files from a CD with Media Player. A variable bit rate does not have a fixed allocation of bits, it uses more when needed and less when possible. Typically, variable bit rate codecs will sound better as they make better use of the bits that are available. Variable bit rate codecs make a lot of sense when files are not being transmitted over band limited channels. For instance, when a collection of songs are stored on a PC the bit rate is only important from the storage point of view and the extra space of the variable bit rate is not a significant factor. Fixed bit rates work better for transmission since the bandwidth required is a know quantity.

Variable bit rate
a technique were the instantaneous bit rate is allowed to fluctuate.

3.2.c Dolby Digital

Also called AC-3, this algorithm differs slightly from other standard techniques in that it encodes the bin coefficients **mantissa** (the decimal portion) and **exponent** differently. It uses the exponent to track the waveforms envelope and allocates the bits in the mantissa based on the envelope and a psychoacoustics model, much like MP3. AC-3 is a 5.1 channel system

Mantissa
the decimal portion of an exponential number representation.

Exponent
the power of ten that scales the mantissa.

which means that it has five discrete full bandwidth channels and one limited bandwidth channel for a sub woofer.

AC-3 is currently the broadcast standard for HDTV and is the most common format for DVD; This makes AC-3 an important codec since in a few years it is quite possible that more audio will be coded in this format than any other.

3.2.d DTS

DTS is an eclectic mixture of compression schemes, some of which are perceptual based and others are not. DTS is currently a competitor to Dolby Digital in the 5.1 arena. The interested reader is referred to their web site for further information. DTS is popular for music based DVDs.

3.2.e Others

There are many other perceptual coders that have appeared and some have even found implementation in significant areas. Advanced Audio Coding (AAC) is an MPEG-4 audio standard which has had contributions from a number of companies including Fraunhofer, AT&T and Sony. It is optimized for low bit rates as would be used in Digital Satellite systems. It is licensed by Dolby Labs.

3.3 Conclusion

In conclusion, I want to make a few comments about compressed audio based on my own personal experiences. First, compressed audio is, for the most part, comparable to uncompressed audio if a good codec is used and a sufficiently high bit rate. It seems unreasonable to me to use anything but the highest bit rate available if one is storing music for playback in a quality Home Theater. Storage is just too cheap to worry about storage requirements. I think that it would be unfair to rank codecs other than to say that good ones do exist (while some are pretty bad). I am <u>not</u> saying that compressed audio <u>is</u> indistinguishable from uncompressed audio because under carefully controlled tests, with well chosen source material, the two can always be distinguished from each other. But the better codecs at the higher bit rates take an extremely sophisticated listener to detect and even then only on some program material. Obviously loss-less encoding is inaudible.

Ripping
the act of converting uncompressed audio into compressed audio.

The point is that compressed audio offers so many advantages that the small degradation in performance is, in my opinion, well worth the trade-off. I have **ripped** (used a perceptual coder to encode a song) my entire CD collection into a central computer on a home network. As a result, the access to my music is unprecedented. Gone are the days of searching through stacks of CD's trying to find the song that I am looking for or the tedious organizing that is required in order to better facilitate a search. I can find and play any song in my collection of 700+ CD's from any room in my house in seconds. With loss-less encoding the original CD is only needed once and it can be put away for good. (I don't encourage illegal ripping of music, it's simply not necessary, and clearly unfair to the artist.) The sound of the ripped file is fine for almost any situation.

CHAPTER 4 ACOUSTICS AND SOUND REPRODUCTION

I once heard George Lucas quoted as saying "Sound is half of the motion picture experience." This has always struck me as exaggerated, but clearly, in his films, sound is certainly an extremely important element. There is no doubt that playing a Lucas film through five "mini-cubes" with an eight inch subwoofer is not going to create the audio effect he intended his audience to experience. In this chapter, I will talk about sound reproduction with the assumption that the goal is to recreate in the home the same sound effect that is intended and experienced in a commercial theater. The first topic at hand will be to present some basics in acoustics.

4.1 Loudspeakers

4.1.a Frequency response

Free Field
a measurement without the influence of boundaries.

Axial Response
the dB SPL found at a field point directly in front of the loudspeaker.

One of the most common and important characteristics of a loudspeaker is its frequency response, which usually refers to its axial **free field** response. The **axial response** is important to the loudspeaker evaluation because this single unique direction, usually normal to the front baffle along the axis of the loudspeaker, gives us the “average” response of the system. It is free field because we usually see the loudspeakers response without the influence of room boundaries etc.—the acoustic field is “free” of boundaries. While the axial free field response is certainly a useful measurement, it is only one of many measurements that are required to characterize a loudspeakers acoustic radiation. And, in fact, it is a very small contributor to the more important power response of the loudspeaker. The singular axial response has the smallest associated area of any of the so-called polar response measurements. It is my feeling that the axial response, or any single directional response, is highly overrated. My prference for loudspeaker layout does not even put the listener on axis of the loudspeaker, further relegating this response to the unimportant category.

Unlike typical electronic components, which have a single input and a single output, a loudspeaker actually has a single input, but a multiplicity of outputs. That’s because loudspeakers emit sound into a full three-dimensional space. While the acoustic field in this space is truly three-dimensional, one of the spatial dimensions is not that complicated. For all practical purposes, the radial dimension is not so important because free field loudspeakers always fall off at the same rate with distance (for a sufficient distance) regardless of frequency or position. (This is primarily, but not completely true for all situations, but is true for our purposes.) The fall-off rate goes as the inverse square of the distance, that is:

$$p(r) \propto \frac{1}{r^2}$$

As a result, the usual practice is to simply fix a radial distance and refer all frequency responses to that distance. We can get the actual dB SPL at any distance by subtracting 6dB for each doubling of distance.

There still remains two spatial dimensions in the acoustic radiation sound field. To fully describe the response of a loudspeaker requires a knowledge of its frequency response in these two remaining spatial directions. For example, a loudspeaker at the center of a sphere would have a radial distance that was the radius of the sphere. The axial direction is that singular direction that the speaker faces in, usually the normal to the baffle face at its center. The two spatial directions are the vertical and the horizontal angles measured away from the axial direction. The polar response of a loudspeaker is the frequency response at points on the sphere other than the axial point. It is most common to see these polar responses given as an angular arcs away from the axial point which is labeled as 0° (see for example Figure 4-2). These plots use the axial frequency response as a reference level. By this I mean that the dB SPL levels shown in the plots specifies only the difference in the response at off-axis points from the response of the system on axis. Sometimes the data is in dB and sometimes it is linear. Clearly, there are an infinite number of these polar responses, and there are a great many ways of showing them. I will return to this discussion in section 4.1.c - Polar Response.

The response of the loudspeaker (its axial frequency response) is usually specified at a fixed distance, usually 1 meter, with a fixed input level, usually 2.83 volts (1 Watt into an 8 Ohm load) and is usually stated as "one Watt @ one meter". The response curve is given in dB SPL. The axial response is the most common, but there is a serious issue with the axial response for many competent speaker designs. In a waveguide system, for example, if the waveguide is round, as is common, then edge diffraction will cause irregularities in the axial response. But most importantly, these irregularities will only exist for a very narrow range of angles about the axial location. If one were to attempt to create a flat axial response in this case (with EQ or crossover design), the power response would be seriously degraded. The axis of a loudspeaker is simply not a good place to be listening from for many loudspeaker designs. However many speakers have such poor off-axis response that the axial one is the only one that is acceptable. The bottom line here is don't judge a speaker on its axial response, this won't tell you much about how it sounds.

When one looks at the frequency response of a typical loudspeaker, they are often surprised by the lack of smoothness of the response. Loudspeakers are never as flat as the electronics that one uses to drive them. A discussion of how flat they should be—what kinds of response irregularities are audible—is a complex subject. I will touch upon this subject, but space constraints won't allow me to cover it in detail. The room affects the source in profound ways, and, in fact, the response of a loudspeaker is virtually dominated by the room characteristics through much of its operating range. What is an audible problem in one room may not be in another. However, there are some characteristics of the loudspeaker/room situation that are stable.

At low frequencies the room response will completely swamp out the loudspeaker response. This occurs in what is called the room's modal region, i.e. the region where the modes or standing waves dominate the response. At some frequency, typically a few hundred Hertz in a typical HT, the modes become so dense that the modal response smooths out and modes, per se, are no longer an issue. Since, in the modal region, the response is dominated by the room I will forestal any discussions about this region to the chapter on room acoustics where I will show how to deal with this issue.

Direct field
the sound field where the source contribution dominates.

Reverberant field
the sound field where the rooms reverberant contribution dominates.

Above the modal region the response seen at the listener's position has two principle components, the **direct response (or direct field)** and the **reverberant response (or reverberant field)**. It is called the direct response because it is the direct first signal arrival that makes up this response. It is free from any room effects since this first arrival has not had time to interact with the room. If the listener position were to be on the axis of the loudspeaker (which is not usually the case), then this response would be the same as the free field axial response. The polar response shows what the direct field response will be at different locations around the loudspeaker.

Power response
the sound power radiated from a source. Basically the average of all sound radiation directions.

The reverb response is a long-term integrated response of the loudspeaker/room system and is dependent on the total sound power response of the loudspeaker (also known as its **power response)**. The rooms absorption characteristics also influence the reverberant response. I recommend placing the axis of the front loudspeakers to point to the approximate center of the listening space, usually just ahead of the listening position. This typically places the listener at an angle away from the axis of somewhere bewteen 15° and 25°. This toe-in helps to eliminate the nearest wall reflection for directional speakers as will be seen in the next chapter. At low frequencies, the period of the waves is actually comparable to the travel time to the listener and, as such, it is not really meaningful to talk about a direct response in this frequency regime. There is only the reverb response.

The low frequency response is strongly affected by both the loudspeaker and the room, but the room tends to have the greater effect. The opposite tends to be true for the higher frequencies. The transition from low to high frequencies will be discussed again in the next chapter on room acoustics.

Coloration
timbre changes that result from frequency response irregularities.

Clearly one wants a flat direct response since there should not be any **coloration** of the all important first arrival signal. By coloration I mean frequency response irregularities cause by resonances or comb filter effects due to reflected signals. The term coloration comes from an analogy with light, where colored light has specific frequency components that are more predominate than others. In general terms the issues about frequency response can be simplified (grossly) as follows:

- Peaks are more audible that dips.
- Audibility of a peak is proportional to the total area under it relative to adjacent frequency bands.
- Gradual changes in response do not create coloration, but do give an overall brighter (rising response) or duller (falling response) sound character.
- Since an overwhelming amount of the sounds in our environment lie in the region of 300–3000 Hz (virtually all speech) our hearing is most acute in this region. Music also tends to be dominated in its energy content in this frequency band. As a result, aberrations in this band are more detrimental than aberrations outside of this band. Being the band of speech this is the most common usage of our hearing mechanism.
- Diffraction effects can be audible in cases where they are not highly notable in the frequency response. These can somtimes be seen in the impulse response, which should be short without any ringing tails or spurious artifacts.

These are simple rules which (unfortunately) have many exceptions, but they do offer some insight into how to evaluate a loudspeaker response curve.

The reverberation response and the loudspeakers power response, are issues that are more pertinent to the topic of room acoustics, so much of that discussion will be left for Chapter 5. I will return to a portion of this discussion when I talk about polar response.

4.1.b Distortion

Distortion in loudspeakers or any audio component in general is a subject that is under much scrutiny at the present. My wife and I have performed numerous studies into the subjective evaluation of distortion which have had some intriguing results. These results are contrary to much of the current thinking on distortion, but the data is now overwhelming. The details of this work are not as important as the more important results. These include:

- There is virtually no correlation between Total Harmonic Distortion (THD) or Intermodulation Distortion (IMD) measurements of a system and the subjective impression of the sound quality of that system. The correlations were weak, but most shockingly they were negative—according to these tests people liked THD distortion. This is actually somewhat true in general - that people prefer some forms of distortion to no distortion.
- Signal based distortion measurements (THD, IMD, MTD), which are based on a purely mathematical formula which does not take into account the characteristics of human hearing, do not hold out much hope for ever being an accurate measure of subjective impression.
- Measures need to be based on the actual nonlinear characteristic of the system and scaled to account for the human hearing system.
- THD, IMD, MTD are symptoms of the root problem—the system nonlinearity. These symptoms manifest themselves in different ways depending on the type of nonlinearity present in the system which causes them to be system dependent. This system dependency prevents these symptoms from being valid guides to the audibility of the underlying distortion.
- Sharp changes in the linearity transfer function, things like hard clipping and crossover distortion, have the most detrimental effect on sound quality. Gradual changes such as soft clipping have a minimal effect.

How does one evaluate a loudspeaker system for its nonlinear performance and what does this performance mean to perceived sound quality? In my opinion, at the present, there is no generally accepted way to measure distortion in loudspeakers. Hopefully, new measurement systems will come on the market which utilize these new insights into the perception of distortion. However there are some concrete facts that arise from the previous discussion.

Currently, there are usually two values given for loudspeakers to indicate distortion performance. Swept THD numbers are often given, but, as I noted above, these have almost nothing to do with the perceived sound quality. Higher orders of harmonics, like the fifth, six, etc. are better indicators, but they are still lacking in a firm quantification of the problem and actually seldom shown—these higher harmonic measurements are seldom repeatable. A second number that is usually provided is power handling capacity and it is often implied that this number represents the maximum input power before distortion, but this is not true. Power handling numbers usually refer to power limitations before catastrophic failure—there is no guarantee that the system will still sound acceptable at this level. Indeed, many loudspeakers are highly nonlinear at these power levels.

Max dB (SPL)
the maximum sound level obtainable by a loudspeaker system

What one wants to know is the maximum SPL level that the system can achieve before it exhibits a poor sound quality due to nonlinearity. A rule of thumb that I recommend is to consider one half the power rating as the maximum input voltage before a sound quality degradation (assume that it is unacceptably distorted at full power). I then use the system sensitivity to calculate the **Max dB (SPL)**. For example, a loudspeaker might be stated as having 84 dB (SPL) (1 W @ 1 m) sensitivity with a 40 W power rating. Using my rule of thumb, I would estimate the Max dB (SPL) at a listener position of 12 ft (4 m) to be:

$$Max\,dB(SPL) = 84\,dB\frac{meter}{Watt} + 10\cdot\log(20)\,Watts - 20\cdot\log(4)\,meter \cong 85dB$$

The above equation is a good example of the usefulness of the dB scale (and my rationale for covering it in Appendix II). Note that for the power factor, I have used 10 log() while for the distance factor, I used 20 log(). This, too, is explained in the appendix.

A slightly larger system might be given as 86 dB and 50 W which yields a Max dB(SPL) in this room of about 88 dB. If the rooms reverberation is considered (if it is somewhat reverberant), then about 3 dB can be added to this number. For a medium reveberation room, 3 dB is about right since 6 dB is the maximum level increase that a very live room would achieve at a typical seating location. This increase would go to zero for dead or open rooms. If one is sure that the power ratings occur at low distortion (but this would be the exception), then another 3 dB can be added. The above example will give about 88–91 Max dB (SPL) and 91–94 Max dB (SPL) for two systems. In later chapters, I will show that this is not nearly enough output for high performance HT applications.

In loudspeakers, sharp changes in the linearity curve do not usually happen at low amplitudes of cone displacement (but they can happen). This means that nonlinear distortion is not usually a significant factor for typical cone displacements in loudspeakers. In a loudspeaker the cone displacement must go up (rapidly) as the frequency goes down for a constant level of sound output. Thus, in a loudspeaker, the most likely source of audible nonlinear distortion results from low frequencies.

The situation gets more complex when one considers that most signals contain both low frequencies and high frequencies. Since the presence of low frequencies causes large cone displacements, these large displacements will also distort the higher frequencies even though they themselves do not cause high cone amplitudes. A loudspeaker is thus limited in its performance by how much excursion the driver can produce before high rates of change appear in the linearity relationship. This maximum excursion number is usually called X-max. Unfortunately, no standard definition for its determination is currently in force, and the market place is sadly abusing the X-max concept. With a known X-max and cone area one could calculate the Max dB(SPL) in a more accurate (albeit slightly more difficult) manner than I did above. However, we cannot usually rely on the published values of X-max as being a good indication of true linear capability.

But the situation is not as simple as just looking at X-max (reliable or not). The audibility of a nonlinearity in a loudspeaker inceases with the loudspeakers bandwidth due to the intermodulation effects mentioned above. This means that a subwoofer would not be a major source of distortion, even with a small X-max, because of its small bandwidth. The biggest concern will be the low frequency transducer in the main speakers. This driver should either be fairly linear over a large excursion or have its low frequency output limited. Attempting to take a small speaker very low in frequency will inevitably lead to serious distortion problems at higher SPL levels.

Distortion mechanisms in a loudspeaker are not always excursion realted, some are current dependent. The current tends to be independent of frequency. These distortion mechanisms are thus not limted to low frequency content, but they are also depnedent on drivers bandwidth. Hence, in a wide bandwidth transducer one must address these current nonlinearities in some manner even if the excursion nonlinearities are under control. The usual means of accomplishing this is to use a shorting ring in the motor structure of the loudspeaker. The ring tends to linearize the current realted nonlinearities. It is usually difficult to find out if a loudspeaker driver has one of these rings or not, but it is an important factor in sound quality.

What should be clear from all of the discussion in this section is that there is a strong relationship between the size of the loudspeaker, its design and its ability to produce high levels of linear output. For a fixed design (in other words two drivers of comparable design), as the cone area goes up, the excursion will go down for a given output. Therefore, for each doubling of cone area the loudspeaker can be expected to produce 6dB more linear output. It should be evident by now that high outputs require large cone areas in the individual loudspeakers. Of course the excursion limitations of the design do come into the picture, i.e. a loudspeaker with more X-max produces more linear output, but X-max is a design trade-off in terms of cost and performance. For equivalent cost and performance one should expect the larger loudspeakers to have a substantially greater linear output. Doubling the cone area is far less expensive to do than doubling the X-max. Often doubling the area increases the efficiency while doubling the excursion decreases it. It is all a trade-off, but it is one that always favors the larger size drivers.

In a HT system which is operated in a small room with a television as its video source and the viewer only a meter or two away, the loudspeaker system is likely to only be played at say 70–80dB output. But a HT with a screen of 12 ft or so, a viewing distance some 12–16ft back, and an acoustic output which is comparable to a commercial theater would be expected to reach 100–110dB of linear output with peak demands to as much as 120dB. This additional 30–50dB of output would require an equivalent area loudspeaker in the larger HT to have from 30–100 times more excursion capability than the smaller system in the smaller room, which simply cannot be achieved—at any cost. A more feasible approach would be to use a larger driver with about four times more excursion and ten times more cone area.

Commercial theater sound levels are always much higher than typical music levels, due to special effects, etc. and typical stereo loudspeakers are not going to be adequate in a good HT. Clearly, small speakers simply will not work in a high performance HT without massive amounts of distortion. The loudspeaker design for a HT is a major issue and only rarely have I seen the loudspeaker capability done properly in a HT. Almost without fail, common hi-fi speakers have been used, but these systems simply do not have the capability to produce the acoustic levels necessary for a good HT.

4.1.c Polar Response

Returning now to our frequency response discussions, I will describe the issues that arise regarding sound radiation in non-axial directions.

Sound radiates from a point source in all directions uniformly. This is the so-called monopole response that has equal acoustic output levels in all directions. Real sources are not point sources; however, and, indeed, we learned from the last section that in a HT the loudspeakers should be fairly large in order to handle the higher output requirements of film. Larger sources do not send sound in all directions equally.

As I said before, the polar response is basically the response of the loudspeaker at points other than the axial direction. The common ways of discussing the polar response is to talk about the vertical and horizontal responses. Typically, this is done by showing the frequency response at some non-axial point or showing the response at a single frequency as the angle away from axial is varied (as in Figure 4-2). A common misunderstanding is that a loudspeaker is fully quantified by its vertical and horizontal polar response. This is not true because these responses quantify only two circles in the spherical space surrounding a loudspeaker. There is no guarantee that things happen predictably in the spaces in between these two circles. My preferred method for looking at polar response is with a contour plot that is called a polar map. Once one gets used to this type of plot, it is quite instructive.

Wavenumber
the number of wavelength per unit distance (see pg. 24).

In order to discuss the radiation characteristics of a loudspeaker, it is useful to introduce an acoustical variable called k, the **wavenumber**. The wavenumber is simply 2π times the inverse of the wavelength. A really useful quantity is the wavenumber times a distance a because this is a dimension-less number that simply describes the number of wavelengths in the distance a. The value ka is the acoustic dimension parameter, for it sets the acoustical size of an object. Mathematically the relationship is

$$ka = \frac{2\pi f}{c} a$$

where *c* is the speed of sound. We can see that *ka* is directly proportional to frequency. The variable *ka* then describes a distance in terms of the number of wavelengths, i.e. *ka=1* is a distance of one wavelength.

If *a* is the radius of the loudspeaker then *ka* is simply half the number of wavelengths across the diameter. Figure 4-1 shows the *ka* value for differ-

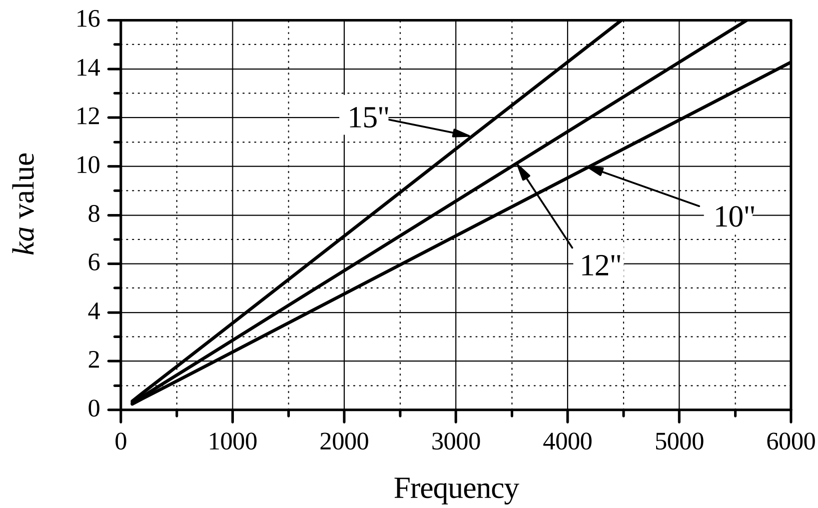


Figure 4-1.
ka value versus frequency for three sized drivers

ent loudspeaker sizes as a function of frequency. The reason that this variable is so convenient is that it turns out that the acoustical properties of acoustic sources depend almost exclusively on the *ka* value. A plot of sound directivity, polar response, versus *ka* is therefore applicable to any loudspeaker, regardless of its size. This is a most convenient feature.

Figure 4-2 shows one of the typical means for showing the polar response. In this curve, the distance of the curve from the center of the circle represents the level of the sound radiated in that direction. The vertical line is normal to the source and originating at its center (its axis), and it is customary to normalize this type of plot to be common at the axial point (the top of the circle in the figure). This figure shows typical polar plots for three values of $ka = 2, 4$ and 8.

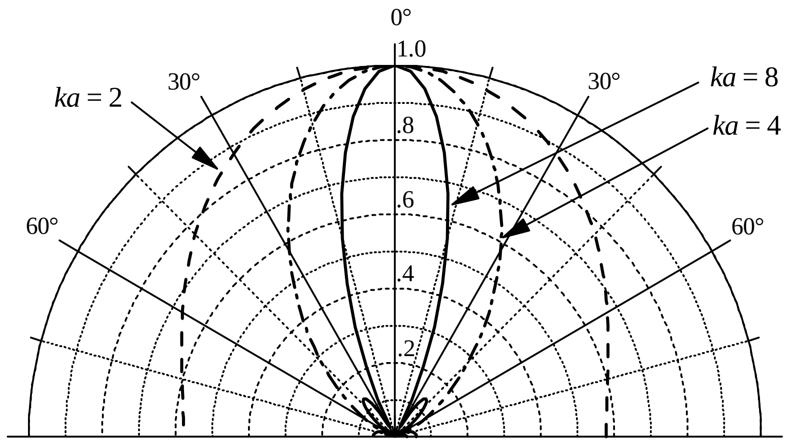


Figure 4-2. Polar response for ka = 2, 4 and 8.

From what I said above, it should be clear that the plots in Figure 4-2 apply to any loudspeaker, but at different frequencies for the different sizes. For example, the polar pattern for *ka* = 4 occurs at 1100 Hz for the 15” loudspeaker, 1400 Hz for the 12” loudspeaker and 1700 Hz for the 10” one. This polar pattern would also apply to a 1” tweeter at about 16 kHz. This plot shows the linear level, it is not in dB, which is different from later plots in this text, which are almost always shown in dB.

It will be useful to show a different kind of polar plot than the one shown above—one that contains a lot more information and, hence, will be more enlightening in coming discussions. Unfortunately, it is a little more complex and harder to follow, but the effort in understanding it will be worthwhile.

A better representation of the polar response can be shown by plotting it as a contour plot. A contour plot is one where the dependent variable, the dB(SPL) in this case, is represented as a color, a gray scale in our case. White represent the highest levels and black the lowest levels. Black regions have less than –21 dB of signal or about 10% of the maximum level (white). Each level change is –3 dB. There are two axis in this type of plot, angle as the y axis, and *ka* as the x axis. A polar plot of dB SPL versus angle and *ka* value for a rigid piston is shown in Figure 4-3. The entire polar response for an axi-symmetric loudspeaker (round) can be shown in a single figure with this kind of plot.

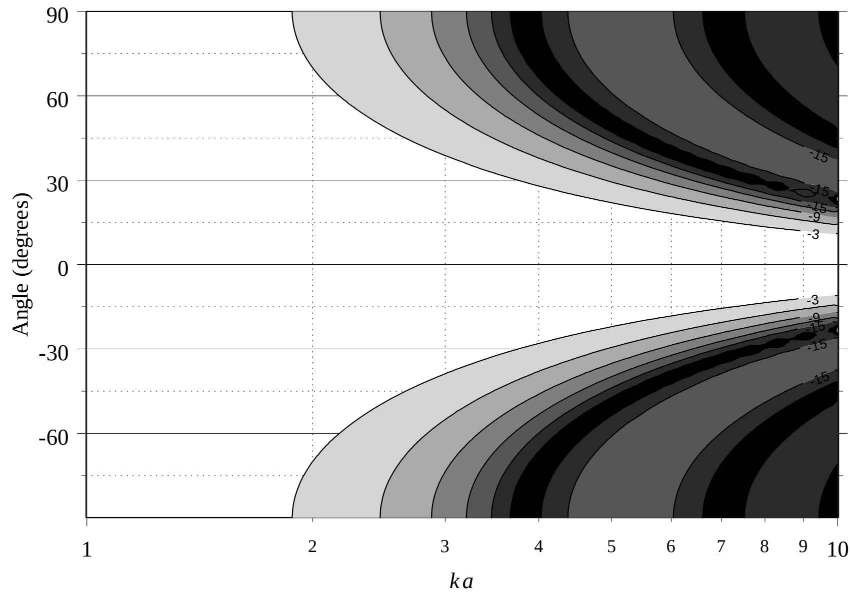


Figure 4-3.
Polar response as a contour plot.

The white space at low values of *ka* corresponds to a completely uniform (monopole) acoustic field. As *ka* goes up, the acoustic field drops to –3 db at the baffle (this plot is for a baffled loudspeaker) when *ka* is about 2.0. From Figure 4-1, we can see that this happens at a frequency of about 500 Hz for a 15” loudspeaker and about 800 Hz for a 10” loudspeaker. The –6 dB points are called the **beam width**. From the figure above we can see that the beam width is about 80° when *ka=3*, from Figure 4-1 on page 63, for a 12” loudspeaker, this occurs at just abovet 1 kHz. The combination of Figure 4-1 and Figure 4-3 can be used to determine the beam-width for any circular transducer at any frequency. Real loudspeakers differ somewhat from these idealized cases, they are usually a little wider in response because of the cone, but for quick reference these curves are fairly accurate.

Beam width
the total angle in the axial direction for which the sound radiation is above 6 dB.

Figure 4-3 shows the polar response only in a single plane, which is fine for a device with axi-symmetry and, in fact, only half of this figure is unique in that case. But for a complex system which does not have this symmetry we must look at several planes to get the whole picture. At the very least, we should look at the horizontal and the vertical planes,

Omni-directional a source that radiates sound equally in all directions.

although, as I said before, this coarse picture can sometimes miss some important flaws in the polar pattern.

Note that below $ka = 2$, the sound radiation is equal in all directions, it is **omni-directional**. This fact leads to the rule of thumb that all sources are omni-directional until they are about the size of a wavelength of the radiated sound. Said another way, objects are acoustically featureless if they are smaller than a wavelength. A convenient number to remember is that the wavelength at 1 kHz is about 1 ft. There are some important implications of this fact. An enclosure or an object is basically shapeless to a sound wave—it will have the same characteristics regardless of its shape, round, square, etc.—below $ka = \pi/2$, or $a = \lambda/2$. At frequencies above the point where its largest dimension is about the size of a wavelength (remember that a is the radius not the diameter), the enclosing object will begin to affect the source in ways which depend on its shape. Below this frequency an object's shape is relatively unimportant. For example, at frequencies below $ka = \pi/2$ ($f = c/4a$), things like cabinet edge diffraction depend almost exclusively on the enclosures volume and not on its shape. There is always diffraction, even at low frequencies, but the amount depends on much larger features than the cabinet edges. Above this frequency, an object's shape has a much larger effect and the baffle edges become major points of diffraction. These effects are completely analogous to radar where features in the reflected object cannot be detected below the wavelength of the radar signal.

Now that I have shown a good way of mapping the polar response, we can move on to one of the most critical features of a loudspeaker design and that is its power response. The power response is the sum, or average, of the sound radiated in all directions as a function of frequency. The power response for the polar map shown in Figure 4-3 is shown in Figure 4-4. Figure 4-4 is obtained from Figure 4-3 by averaging the values in the polar map along vertical lines—ka values. Figure 4-4 then plots these values versus ka, giving a simplified picture of the total radiated sound from the more specific picture shown in the polar map. This method works fine for a source that is axi-symmetric but not for arbitrary sources. In the later case, we must look at many such polar plots around the source.

The power response is seen to fall dramatically with frequency as a result of the increasing directivity of the source. An examination of

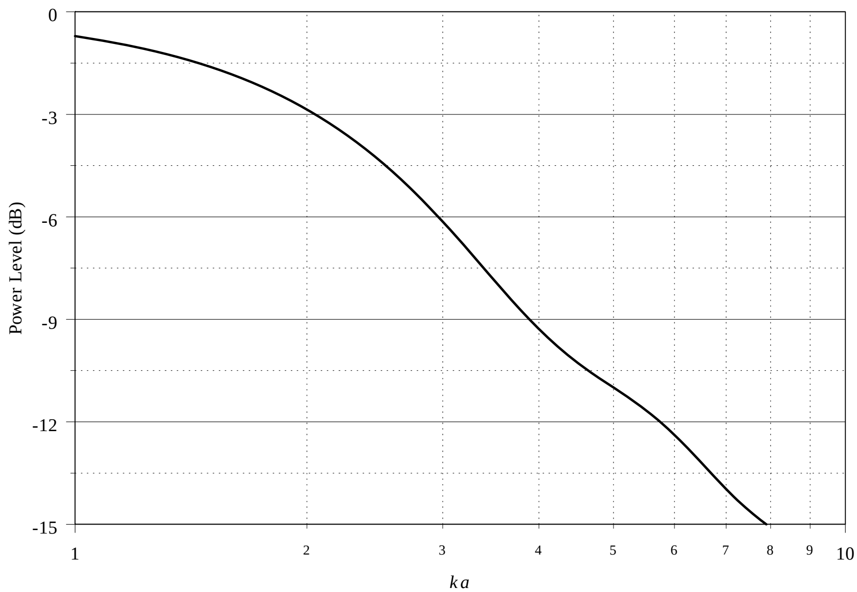


Figure 4-4. Power response for an ideal piston.

Figure 4-3 and Figure 4-4 shows that a typical loudspeaker that has a basically flat axial response will have a falling power response. This means that the direct, first arrival sound can be flat, which is a good thing, but unfortunately, the reverberant field (See "The Direct and Reverberant Fields" on pg. 90.) will not be flat, which is not so good. To make matters worse, sound equalization in rooms is almost always done using a steady state or reverberant sound measurement (see Appendix IV). Equalization of a typical loudspeaker in a real room using steady state measurements can cause the axial (and hence direct) sound to be seriously degraded. The solution to this problem is to have both a flat axial response as well as a flat power response. But, as we shall see, this is easier said than done.

The usual means for achieving the above simultaneously flat sound fields is to design the system to have multiple drivers, thereby obtaining a better match between the axial response and the power response as the system steps through the different drivers. There are some substantial issues associated with the use of multiple drivers, and I will discuss these in the next section.

4.1.d Loudspeaker Systems

It might be advisable at this point to define some trems, even if these terms are widely know. In this text, a loudspeaker means the system, it includes the drivers, enclosuer and crossover(s). A driver is the indvidual transducer element of which there can be many in a loudspeaker.

I have shown several problems with the use of a single driver to cover the entire audio bandwidth. Lets review these:

- Low frequency cone excursion can cause distortion of all frequencies in the pass band of the driver. This is true even though the nonlinearity is limited to low frequencies.
- The power response of a single piston type driver is not aligned with its axial response, the two things vary at different rates. This means that they cannot be corrected simultaneously with electronic means. They are characteristics of the acoustical system.
- Drivers cannot cover the entire audio bandwidth without significant resonances occurring at the uper edge of its bandwidth. These resonances are clearly audible and must be controlled or avoided. Controlling them is difficult to impossible, but they can be avoided by transferring the signal to another, usually smaller, loudspeaker when these resonances start to occur.

For these reasons (and a few more not discussed) good loudspeaker systems have multiple drivers with each driver covering a specific bandwidth for which it is optimized. The principle design problem for a loudspeaker system then is concerned with the optimal transition from one driver to the next. This is not always an easy task and it is this feature which tends to make or break a good design.

Cone break-up occurs when the loudspeaker cone no longer moves as a unit.

To start out this discussion, let me describe a rather idealistic system in which there are two drivers which are at exactly the same point is space but are of different sizes and cover different frequency bands. Let the radius of the smaller one be 1/4 that of the larger one, say a 12” loudspeaker and a 3” loudspeaker, not an uncommon situation. Any absolute sizes work, so long as the ratios (the *ka* values) stay the same. The 12” loudspeaker would likely start to have **cone break-up** (the term used to imply the start of a multiplicity of resonances) at about 1000 Hz or a *ka* value of about 3.0. Let me pick the crossover point to be just below where this breakup begins to occur, about 900 Hz. At 900 Hz the smaller speaker would be at just under *ka*=.75, so the smaller speaker would be omnidirectional at this frequency. The polar response map would look like that shown in Figure 4-5 below. The crossover point is obvious at about 900 Hz, and the woofer's directivity can clearly be seen below this point. The power response calculated for this setup is shown in Figure 4-6 on page 70. This power response is a major problem because the reverberant sound is more dependent on the power response than the axial response. As I noted before the axial response has almost no effect on the power response because it represented by a very small spatial area. The reverberant energy will be lacking in energy at the crossover point. The impact of this flaw depends on the characteristics of the room, and I will return to this discussion in the next Chapter. Also shown in this figure are the pseudo-power responses, power calculations out to only 60° and 30°, which are often stated as being the more important angles since they more heavily weight the forward direction. These curves show that even the narrower angular consideration does not alleviate, the problem that I noted above.

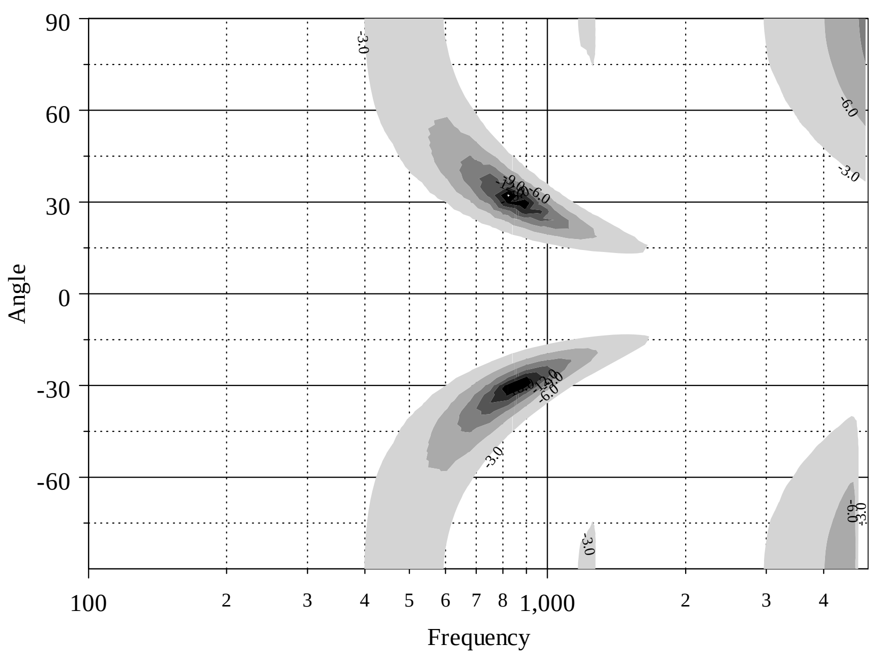


Figure 4-5.
Polar map for an ideal two driver system

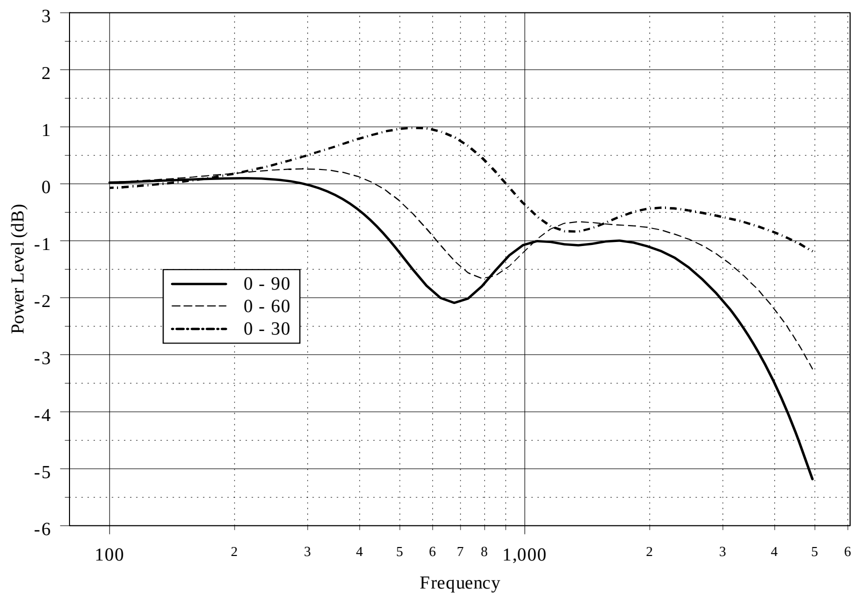


Figure 4-6. Power response for a typical two speaker system

I reiterate that the situation shown here is ideal in that there is no spacing between the two drivers in the calculation. The hole in the power response is due to the crossover phase shifts and ot the loudspeaker spacing. Since this is usually not physically possible, there will be another factor to the polar response which usually aggravates, but always affects the crossover problem. Without going into too much more detail, it is sufficient to say that crossovers between drivers have serious problems to overcome and their design tends to be the Achille's heal of the loudspeaker system design. I would like to point out, however, that no crossover design is magical in that it can make these problem go away. Indeed, being predominately acoustical in nature, no set of electronic filter components can eliminate it. The crossover can be made to minimize the acoustic problems, but it can never be made to correct them.

Note also in Figure 4-5 on page 69 that the axial response is basically flat, an ideal situation if one only looks at the axial response. Clearly doing the crossover design to yield a perfectly flat axial response does not necessarily reult in an ideal polar/power response and certainly not an ideal overall situation. A flat axial response is no proof of a good loudspeaker.

In the above situation, I could lower the crossover frequency to 500–600 Hz, which would substantially improve the power response at the crossover point, but it would further degrade it at higher frequencies. This would then require yet another crossover and the power response aberration associated with it. Add to this the fact that taking a 3” driver down to 500 Hz would cause its cone excursion to rise dramatically and its nonlinear distortion to increase in proportion. The loudspeaker system design problem is multi-faceted and not easily solved!

The solution that is often found in the marketplace is to use two small speakers—say 6" or 8" woofer and a 1" tweeter. Properly designed, a system like this can nearly achieve both a flat axial and power response, but it is inherently omnidirectional. I will show in the chapter on room acoustics that omnidirectional is not really what we want in a HT. There is also the problem with the Max dB(SPL) that I talked about earlier. This simple small speaker design, while being reasonably good in its frequency domain characteristics, is never going to achieve the high outputs required in a good HT. Making the sources larger to handle the required output, leads directly back to the problems associated with directivity that I have just described. The situation seems to be getting away from us.

Waveguide
a “horn” like device, but one which has been designed for directivity control over efficiency.

Constant Directivity
a loudspeaker or system whose polar pattern is constant with frequency.

The solution that I recommend (among the others that I will discuss in this text) is the same one that is used in commercial theaters. That is the use of **waveguide** devices (traditionally called a horn). The advantage that these devices have is that, if properly designed, they can achieve **constant directivity** (CD) with frequency. Of course, no system can have constant directivity at all frequencies, there must be some widening at the lower frequencies (I will later show why this might actually be desirable), but it is possible to have the coverage angle decrease up to a selected frequency and to remain at that coverage angle through the rest of the transducers bandwidth. An example of the polar map for a CD waveguide is shown in Figure 4-7 on page 72. This figure shows that the polar pattern, beamwidth, falls to about 60° at about 1 kHz (–6 dB at ±30°), but remains constant from that frequency up. The downside of this highly desirable characteristic is that to do this wonderful job of beam width control the devices have to be large. By now, the reader should realize that large speakers may be the only option in a high performance HT.

In the CD waveguide polar map shown in Figure 4-7, it should be appar-

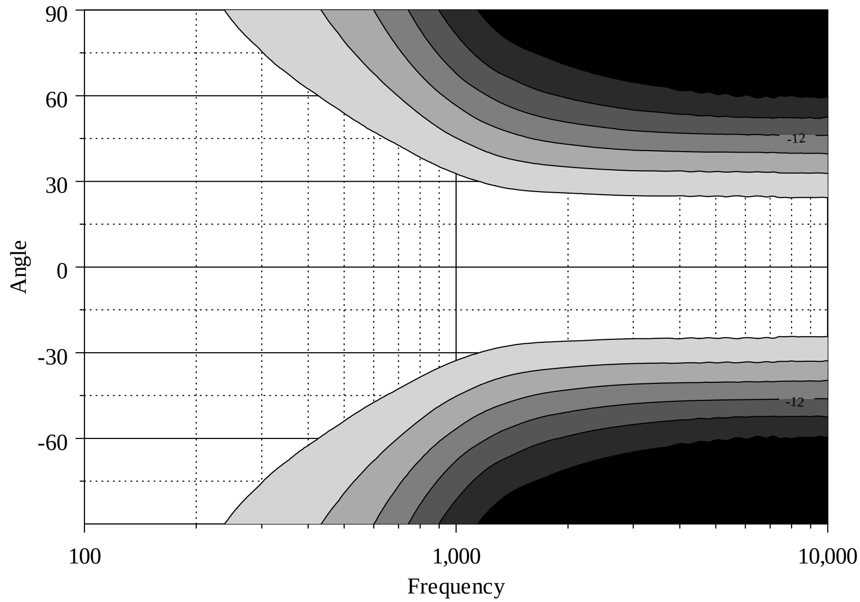


Figure 4-7. Polar map for a CD device.

ent that the power response would be flat from 1kHz up, meaning that the axial response and the power response are tracking one another—which is the ideal. (Unlike the piston source that was talked about before.) This is the reason why the vast majority of commercial theater installation use this technology. Not only does the waveguide produce higher output with low distortion, and have an axial response that is the same as the power response, but it can direct the acoustic energy to the specific audience location—it has a narrow directivity, i.e. it focuses the sound in a specific direction. A waveguide does not send precious sound energy in all directions around the room. It sends the sound in very precise and controllable directions, which is a very useful feature as well shall see.

What is not apparent from Figure 4-7 is that CD devices always require equalization. This figure has the response normalized to the axial level—meaning that the axial response is always set to 1.0 (0dB) at each frequency and the off axis results are adjusted accordingly. This is standard practice in polar plots. In a real device, the axial response actually has to fall with frequency, because the cone velocity is falling above resonance.

But the axial response now does this in exactly the same manner as the power response. The detailed physical reasons for effect can be found elsewhere[1]. The equalization can be done in the crossover or electronically. The lower output required at the lower frequencies (for it is the high frequencies that must be boosted) in a waveguide system is a real advantage to the driver of the waveguide. The driver for these devices usually have a negligible excursion requirement due to the dramatic increase in efficiency that results from the use of the waveguide. Studies have shown that the nonlinear distortion in a waveguide driver are virtually inaudible[2]. Most waveguides are driven by drivers with diaphragms that are larger than the waveguide throat and in this case the drivers output has to be compressed to fit through the waveguide. These kinds of drivers are called compression drivers. Compression drivers are virtually always thermally limited (heat dissipation), not excursion limited as in most piston drivers. If properly designed, waveguides can be made to have extremely high output with very low distortion.

It is also possible to use a waveguide without a compression driver, but this does tend to make them much larger. And size is usually the single biggest problem with the use of a waveguide in the loudspeaker system. A significant lowering of cone excursion requirements can usually be achieved with a waveguide. This comes about because of the increase in the directional properties of the waveguide when compared to the unmodified piston source.

Horns have historically had a bad name becasue of several problem inherent in their design. The first has to do with the way in which older horns controlled the directivity. More modern anaysis of these devices has led to a much improved approach to directivity control that here-to-for obtainable (see [1] below for more details). Further developments have created a waveguide with a sound quality that is on par with the very best tweeters available in the markeplace today. The combination of controlled directivity and very high output capability with very low audible distortion makes the waveguide an unbeatable device for use in a home theater.

1. See E.R. Geddes, *Audio Transducers*, GedLee Publishing, Novi, MI 2002.
2. See Lee and Geddes, JAES, January, 2006

Figure 4-7 on page 72 shows only the high frequency device (the waveguide and driver) alone, and a system would need a woofer for the lower frequencies. It is not too difficult to see that a woofer could be selected which would mate to this polar response almost flawlessly. The only problem to consider is the fact that the woofer and waveguide are not usually at the same location (although this is sometimes done, it is not without its problems), so there will be crossover anomalies to deal with just as with the situation described above. But, owing to the narrower directivity at the crossover point, the power response aberrations are usually not as severe. The use of two woofers can help to resolve this spatial displacement problem by placing one on each side of the waveguide, resulting in an acoustic center for both the upper and lower driver systems that are at the same location. The precise design details are too complex to elaborate on in this text and these systems tend to be on the very high end of the price spectrum.

For a far more comprehensive look at the technologies that I describe in this section, I refer the interested reader to my book *Audio Transducers*, which describes them in excruciating detail.

4.2 Loudspeaker Selection

I will conclude this chapter by saying that the selection of the loudspeakers for a HT is a daunting task. It is one of the most difficult choices that has to be made, and it will be one of the biggest expenses. I have not really given a set of hard and fast rules for selecting loudspeakers simply because one does not exist. I have attempted to show a roadmap that can be used to find the best solution in a particular situation. In every HT that I have done, the loudspeaker selection has been very challenging. At best, I can give you some suggestions as to how I would approach the problem.

The first thing that I am going to say is that the normal mode of loudspeaker selection simply won't work here. Going to the audio stores and listening to loudspeakers is likely to produce more confusion than clarity of choice. That's because the evaluation is not really meaningful. The rooms that these speakers are in are usually not HTs, the speakers are seldom set up where they should be and virtually never can you listen at the kind of output levels that they will experience in practice. If you are arraying systems, then this is altogether a different animal and simple single speaker evaluation is probably not relevant. Go ahead and ask your local retailers to stack four loudspeakers together so that you can listen to them and see what they say.

I suggest that you start out doing some serious research. Use your seating location, the speaker's sensitivity and power handling and determine if one speaker will achieve your goals. If one won't do it, then how many would you need to buy—for each channel—to be able to meet the challenge? Remember that each time you double the number of speakers the MaxSPL goes up by about 6dB. Let me use my previous example where the MaxSPL of the smaller speaker was 90dB. If I want to achieve even the lower end of a SPL goal of 100dB, then I would need four of these for each channel. I would need an incredible 8 of them to achieve 108dB which is still not up to a commercial theaters target output. Quickly, we can see that the sheer cost of achieving the goal makes the choice of this speaker untenable. We don't have to listen to them to know that they will have serious shortcomings in our application.

Once you have found a speaker or speakers that do meet the SPL target, see if you can determine the polar/power response either from the specs or the curves given in this chapter. You are looking for something like a 70°–90° horizontal beam width and a 30°–50° vertical one. Anything in this range is fine. Wider than this can usually be accommodated, although the rooms sound quality might be diminished, but narrower than this may turn out to be a real problem.

Keep in mind that if you are required to use multiple speakers per channel, then they need to be placed close together. Some speakers have front baffles that are not flat and, as such, they would not combine very well since the final baffle would then have a sort of a complex terrain which could have significant diffraction problems for the multiple sources. If you are going to combine speakers to get sufficient output, look for ones with flat front baffles and removable grills. Otherwise they won't combine very well.

Finally, if you find yourself having four or more speakers systems per channel, for heaven's sake, consider jumping up to a professional cinema system. This can often be the most cost effective approach, it will have much more output power and tend to be far more reliable than a large array of consumer speakers. I always recommend that my clients go this route and those that take my advice have never been disappointed. It's what I did in my HT.

The down side of the pro units is usually their appearance, which is not a factor if they are behind a screen. This is what I do and what I recommend.

4.3 Conclusion

This has been an important chapter because without good speakers one cannot achieve good sound and without good sound one is missing a major aspect of the theater experience. Good sound does not come easily, cheaply or in small packages (regardless of the marketing hype to the contrary).

Unfortunately, good information about the polar response of a loudspeakers is usually hard to come by. Odds are, if the manufacturers is not talking about the directivity response of the system, then it is either bad or it's omnidirectional—as all small speakers will be. As I have shown in this chapter, one can usually estimate what the directional response is from the basic design because the polar response of piston loudspeakers is highly predictable. I have shown how to do this estimation and how to use what you have learned to yield a better loudspeaker system by a simple array of individual systems. These techniques will work in some cases but not all.

Larger HTs will require larger systems that would typically be designed for a commercial theater setting. Commercial theater speakers tend to be large, have high outputs with low distortion, and are usually designed to have a controlled, fairly narrow directivity. These are the same criteria that should be targeted in your HT. If you use small speakers in your HT, then you should expect to have a small sound in your theater.

CHAPTER 5 ROOM ACOUSTICS

Room acoustics is an complex subject on which numerous books have been written. However, virtually none of them apply to small rooms. There simply has not been much demand for an understanding of the sound in small spaces. Automobiles are the single most common listening room in the world, but serious listening is just not a high priority. Even in the automobile little work has been done on small room acoustics. In my years working for auto related companies I have both studied and contributed to the state-of-the-art in small room acoustics. Large rooms have large budgets associated with them, and large budgets drive extensive study. The large room acoustic problems are well understood and it is unfortunate that so little of it is applicable to small rooms.

In this chapter, I will talk about the basics of what is known about small rooms, and I will pay particular attention to their differences from large venues. I will not be able to cover this subject in detail (it is a good topic for another book). I will provide you with a reasonable background and some specifics to pay attention to. The detailed theory behind this background information will have to remain transparent for now.

First, however, we need some basics in acoustics.

5.1 Acoustics, Waves and Measurements

Compression Waves waves whose propagating component is the instantaneous pressure or compression.

Acoustics is the study of the propagation of a perturbation of a medium, which can be air, water, structures, etc. By perturbation, I mean a variation in or modulation of some parameter of the medium—like its static pressure or its stress. In air it is the pressure fluctuations about atmospheric pressure that propagates as a sound wave. Acoustic waves in air act much like waves on the surface of water; however, they are different in that acoustic waves are changes in pressure, or **compression waves**, while ocean waves are changes in the height of the surface. For the most part, in the case of ocean waves, the fluid is not compressed. These two types of waves do, however, act quite similarly if we consider that the wave height on the water is analogous to the pressure in the sound wave. One of the more enlightening exercises that one can do to understand wave motion is to experiment (play) with a ripple tank. Water in a shallow tray acts almost exactly like the acoustic waves in a room with the same shape as the tray. If the tray has a glass bottom and it is lit from below (the reverse is also possible) then the waves become visible. My Master's thesis project involved a ripple tank where I had created a wave genertor from an old loudspeaker by gluing a paddle to the voice coil. I could drive the system at any frequency that I wanted (within limits). I learned a great deal about acoustics by watching the propagation of these fascinating waves. Even today, I can't help but throw pebbles into bodies of water to watch the waves reflect off of and diffract around nearby objects. Sound absorption can be simulated in a ripple tank by gradually reducing the water depth— just like waves on a beach, where the waves are almost completely absorbed.

SPL Sound Pressure Level, the RMS value of sound waves in air.

In acoustics, the measure of sound amplitude that we will use is called the **Sound Pressure Level (SPL)**. SPL is a log, dB, quantity and is noted as dB(SPL). The reference level for dB(SPL) is .000002 Nt/m², meaning that

$$dB\ SPL = 20 * \log\left(\frac{p}{.00002}\right)$$

hence in dB(SPL) an acoustic pressure signal whose RMS pressure variation is .02 Nt/m² is

$$dB\ SPL = 20 * \log\left(\frac{.02\,\frac{Nt}{m^2}}{.00002\,\frac{Nt}{m^2}}\right) = 20 * \log(1000) = 60 dB\ SPL$$

Calibrated
yields a precise number for a given measurement—this number is should be traceable to the National Bureau of Standards.

A 66 dB(SPL) signal has a .04 Nt/m² pressure fluctuation.

The sensor that is normally used to measure acoustic waves is a microphone. The microphone is placed at a point in the sound field and the pressure that exists at the diaphragm creates a signal, in volts, inside of the microphone electronics. A **calibrated** microphone has a precisely known relationship between the pressure signal and the voltage signal such that it can be used to measure absolute pressures. An uncalibrated microphone does not have a known relationship, and cannot make absolute measurements, but can still be quite useful for relative measurements. Any uncalibrated microphone can be calibrated if an acoustic signal of known pressure is applied to the microphone. In a good microphone, this calibration will hold over large environmental variations (temperature, static pressure, humidity) while a less expensive unit will often vary considerably over these same conditions.

Today, most common measurement systems use sampled signals, i.e. they are digital. In fact, it is probably true that the only readily available types of measurement equipment on the market today are digital. A significant advantage of the digital signal is that we can use the FFT algorithm that we talked about in Chapter 1 to find the signal's spectrum. Some analog sound level meters are available in the marketplace, and analog appears to work fine if one is not interested in any spectral information. However, if spectral information is desired, then digital techniques quickly become the dominant ones.

Acoustic measurements for our uses are of three principle varieties.

- To measure the acoustic response of a loudspeaker, usually exclusive of the environment.
- To measure the acoustic sound field in a room, usually inclusive of the response of the loudspeaker.
- To measure the noise level in a space.

The first two are closely related measures since room response and loudspeaker response are tightly coupled. The first measure is by far the most common, but, as we shall see, they are usually inadequate as an indicator of how the second measure will come out. Of course, what we want is a correct room response and the usual inadequacy of typical loudspeaker measurements makes this problematic. To make matters worse, the second measurement is difficult to do correctly.

The third measurement above is not always necessary since the presence of noise is usually adequately quantified by a simple subjective evaluation. However, in some cases, noise measurements need to be taken in order to do an adequate job in a complicated situation. I will get back to this topic in Chapter 6.

In this chapter, I will be concerned with the second measurement in the above list. And in the next chapter I will get into the first measurement. Some details of acoustic measurements have been relegated to the appendices, so that I can dive directly into more interesting subjects. Readers interested in more detail about measurements should review Appendix IV - Acoustic Measurements.

5.2 The Small Room

5.2.a Differences From a Large Room

RT_{60}
the Reverberation Time to reach 60dB below the excitation level.

In room acoustics, there appears to be a single number which we all believe that we understand and that is the reverberation time. This is the time it takes for the steady state sound level in a room to drop by 60dB and it is denoted by **RT_{60}**. Many acousticians feel that the –60dB level is far too low and that we can really only perceive sound decay down to about a 20 to 30dB reduction. To me this is quite plausible, but in a small room it is also irrelevant. RT_{60} times in small rooms are simply too short to be of significance or to even measure for that matter. The time required to drop to only 20dB is simply miniscule. Examining why these reverberation times are so low is enlightening.

An acoustic space is primarily characterized by its volume and its surface area—all spaces have these two parameters clearly defined. As a space gets larger its volume goes up more rapidly than its surface area. This is easy to see for a sphere where the volume goes up as the third power of the radius while the surface area only goes up to the second power. As a result, the ratio of the volume to the surface area goes up linearly with the radius and a large space will have a relatively small surface area when compared to a small space. Since virtually all sound absorption occurs at the boundaries of an enclosure, sound absorbing material placed on the enclosure surfaces is much more effective in a small room than a large one. This can also be seen by considering what is called the **mean free path**—the average distance that a sound wave travels before it strikes a surface. Clearly, the mean free path for a large room is ten to a hundred times greater for the large room than the small room. For example, a patch of sound absorbing material placed on the wall in a small room is "seen" by a sound wave ten to a hundred times as often as this same patch in the larger room. Even if the entire room is covered with this material, it will still be seen ten to a hundred times more often in the smaller room than the larger one. As a general consequence of these facts we can say that small rooms are naturally more dead (low reverberation) than larger rooms since some level of sound absorption is always present and it works to dampen the small room much more effectively.

Mean free path
the average distance that a sound wave travels before it meets with a surface.

The mean free path also enters into the situation when we consider the human hearing system. As I stated in Chapter 2, the human hearing mechanism integrates over a time period of about 10–20 ms. The mean free path is usually shorter than this time in a small room. It is virtually impossible to perceive a clear direct sound in small rooms (unless they are **anechoic** chambers). This simple fact accounts for a major portion of the difference between small and large room acoustics. There have been virtually no psychoacoustic studies of sound perception on this short of a scale of time delay. In addition, the fact that there is virtually no chance of receiving a clear direct sound results in some significant implications for small rooms.

Anechoic
a room that is free of any sound refections.

With the lack of a perceivable direct sound component from the loudspeakers, it is important to understand the implications of the steady state sound in rooms. But, in order to understand the steady state, I first need to describe the modal characteristics of a space.

5.2.b Room Modes

Mode
a resonance, but more specifically the shape of the vibrations at that resonance.

It is probably common knowledge that rooms have **modes**—also called resonances. But, I think that there is a real misconception about what the modal behavior of a room means. Figure 5-1 on page 82 shows the simulated frequency response for a room with dimensions of 3 m by 4.2 m by

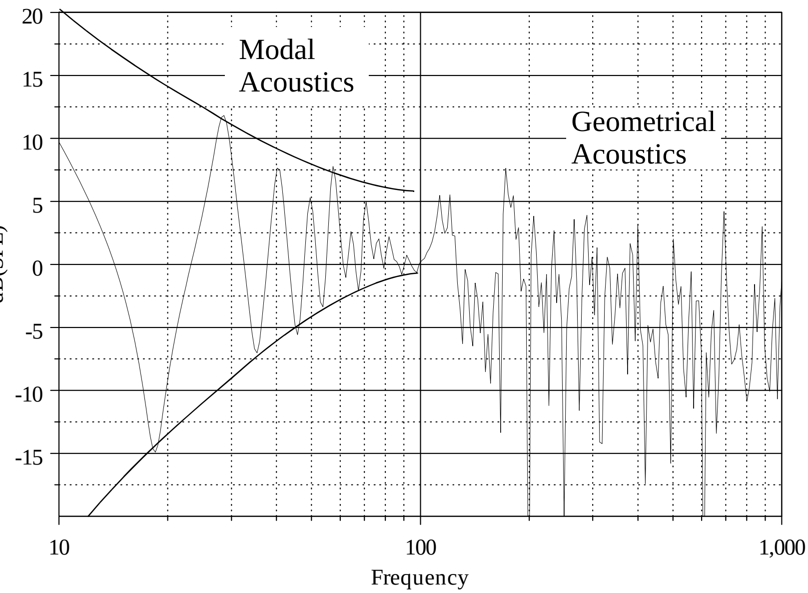


Figure 5-1. Frequency response of a small room in dB(SPL).

6m, a total volume of $76\,m^3$, which is a large HT, but an extremely small room from a commercial theater standpoint. This room is tightly sealed with a light amount of damping on the walls. The first mode (excluding the static pressure mode at 0Hz) occurs at about 28Hz, the next mode occurs at about 40Hz. As the excitation frequency goes up, the modes get closer and closer in frequency—the **modal density** (the number of modes in a fixed frequency band) increases. Note that the fluctuation of the response curve—the "variance"—decreases up to about 100Hz. This decrease occurs because the modes carry and sustain the energy in a room, and, the more of them that there are, the smoother the response will be. This holds true only to about 150Hz after which we see a completely different characteristic evolving. Note also that the mean level is approximately constant up to about 150Hz after which it appears to drop in level by about 3-5dB.

Modal density
the number of modes in a given frequency range.

The region above about 150Hz in this room is called the **geometrical acoustics** region. In this frequency region sound travels in a ray like motion, like light, since there are enough modes to carry sound propagation in any direction. Below the 150Hz point there is a distinct modal characteristic to the room. Modal behavior only occurs at these lowest frequencies. Of note is the fact that this transition frequency (called the **Schroeder Frequency**

Geometrical Acoustics
The frequency range in a room where the sound waves travels as rays—like light.

Schroeder Frequency
the frequency at which a room transitions from statistical to geometrical.

(f_s) after the German acoustician Manfred Schroeder) moves lower in frequency for larger rooms. In fact, virtually all commercial theaters have a Schroeder frequency which is low enough that any modal effects are all below the audio bandwidth. Commercial theaters and auditoriums simply do not have to consider the modal region of a rooms acoustic sound field.

5.2.c Modal Chaos

Statistical acoustics the region where the sound field can only be discussed statistically.

It is important to understand the nature of the sound field in the geometrical acoustics region, above the Schroeder frequency, as this is most of the audio band. This region is sometimes called the **statistical acoustics** region for reasons that I will now elaborate on.

Note in Figure 5-1, that above f_s, the nature of the frequency response curve looks pretty much constant, independent of frequency. It has numerous peaks and dips, but these do not appear to get narrower and closer together with increasing frequency as the response below 100 Hz does. In order to examine this feature in more detail, consider the frequency response curve in the previous figure zoomed in to a narrow region between say 700 and 750 Hz (the actual frequency range is not important). I have shown this zoomed in response in Figure 5-2. Note that there are three curves in this figure. These three curves are frequency response curves for the same room, but with a minor perturbation of the wave speed or alternately the measurement location (they amount to basically the same thing). This kind of wave speed perturbation will almost always occur in a real room, since the amount of speed change is so small. A slight temperature change or a small air current or a person in the room shifting listening position slightly, almost anything would cause perturbations of a magnitude sufficient to cause the response variations shown in Figure 5-2. These three curves appear to be completely different, and they are in detail, but as it turns out their statistical properties are very consistent. They all have the same mean and the same standard deviation; statistically each one of these curves can be thought of as a single sample from the same stable (stochastic) random variable.

The statistical properties of the frequency response curve in rooms are well known. The distribution of dB levels is not symmetric about the mean line, the curve lies above the mean 33% of the time and below it 66% of the time. It lies 66% of the time in a dB band about the mean level that is plus

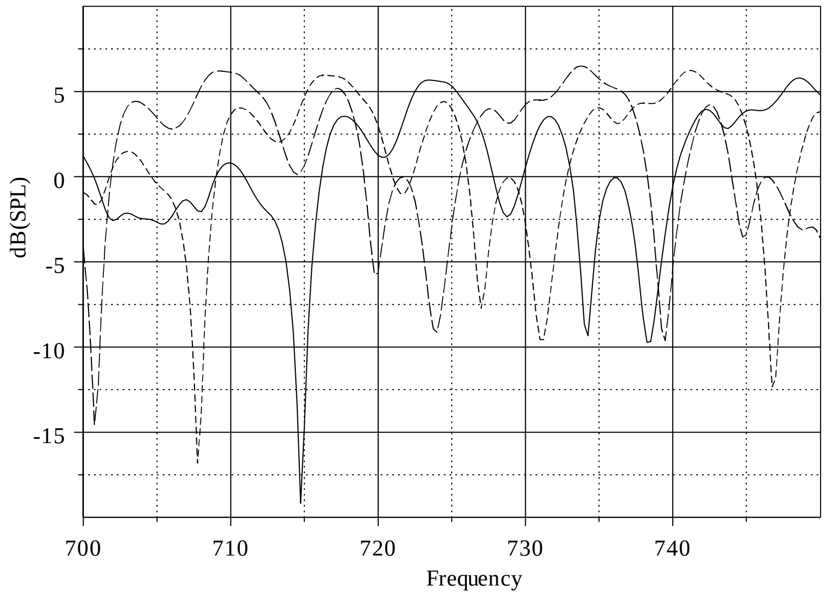


Figure 5-2. Detail of the frequency response in a room at higher frequencies—in dB(SPL).

6 dB and minus 11 dB. The really important thing to remember here is not the statistical values themselves, but the fact that the frequency response curve above f_s, is not a deterministic quantity. A single point microphone measurement of a pure tone that is intended to measure the mean response of the room will actually fluctuate about this mean by the amounts that I have noted. Stated another way, if one attempts to measure the mean level of a tone in a room, it will have an expected error of +6 and –11 dB—hardly a useful measurement. Basically, a single point microphone measurement will not result in the actual room response, it is merely one sample of the room's statistical response.

This can be seen in practice on a spectrum analyzer. Excite the room with a pure tone and watch how the level on the analyzer varies in time. This is the effect of these very small perturbations of the frequency response. The wider the bandwidth of the analysis the smaller the level variations will be, but of course, the poorer the resolution of the response.

These considerations are important to note, but I should also point out another important facet of this discussion. These statistics only apply to the steady state response in the room, not to the direct response of the loudspeaker. In many small rooms the direct response from the loudspeaker is usually quite short since the reflections often arrive within a few milliseconds and the room reaches a steady state in little over this time period. The point here is that in a small room there can only be statistical measurements as nothing else makes much sense or is meaningful. This has profound implications to measurements of a small room's acoustics. (Later I will show how it is possible to make meaningful measurements in a small room at very short times, but this takes special processing.)

The standard way of dealing with this issue is the same one that is used for the investigation of any random variable, and that is to take a number of samples and average them, thereby reducing the variability. It turns out that one can achieve this averaging effect in two different ways. The first is to average over frequency, and, if the measurement bandwidth is of sufficient size, then the measurement variability will drop down to something more useful (+6,–11dB variations are not really very useful). However, to get down to a ±2dB measurement would require a bandwidth of over an octave. This is generally felt to be too wide, since it is much wider than the critical bandwidth of the ear, which is approximately 1/3 of an octave. In order to get better spectrum resolution, one that is consistent with the resolving power of the human ear, we need to supplement frequency averaging with another technique for reducing the measurement variability.

The second technique for reducing the statistical variance is to take several measurements over different spatial positions. This, too, will lower the variance if the measurements are sufficiently far apart in space. The rule of thumb is that each measurement should be at least $\lambda/4$ apart at the lowest frequency of interest. A good estimate of this would be about a foot apart. Roughly speaking, a measurement of the steady state sound field in a room will be about ±1 dB of the actual mean level if four 1/3 octave bandwidth measurements are made no closer together than about one foot. These dB levels must then be averaged together in dB (this is important since time averaging of a spatially moved microphone will not be correct). Narrower bandwidths will require more spatial points, and wider bandwidths will allow for fewer spatial points for a given resolution.

An example of these techniques is shown in Figure 5-3. This is a plot of the same room as shown in Figure 5-1, but with a frequency average of one-third octave at three points separated by one foot. This figure clearly shows that large variations in the response, even when frequency averaging

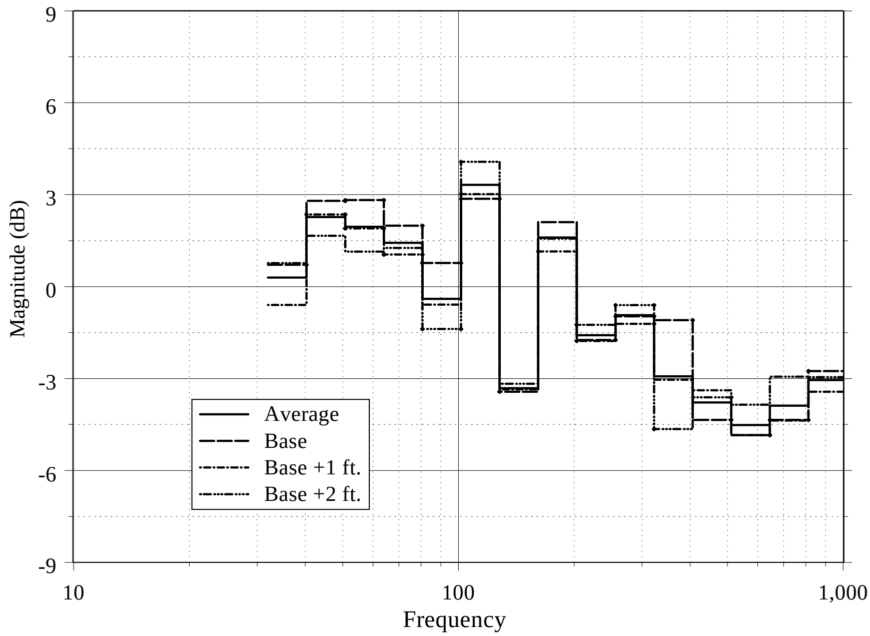


Figure 5-3. One third octave frequency responses for a room at three nearby points

is employed, can occur for spatially close points. Note that the 350 Hz band has more than 3 dB of variation in the response. The solid line is the spatial average of the three independent measurements, and it is expected to be accurate to approximately ±1 dB. The large variations between 100 and 200 Hz. are real and would need to be dealt with, but note how difficult this would be to correct. There are no other problems that stand out in this example. The low end looks quite reasonable in these curves, but at these low frequencies, one-third octave filters are not sharp enough to distinguish the modal variations in the response.

Another noteworthy feature is the much lower response at the higher frequencies. The response above 300 Hz is more than 3 dB lower than below 100 Hz. In the plots shown here, the frequency only goes up to 1 kHz. This is because of the huge calculation times required to go any higher. This is a modal calculation in which there are 27,000 modes in the summation. In a real measurement one would normally take the frequency up to about 10 kHz, which would be hundreds of thousands of modes.

I don't think that many people actually make acoustical measurements of this kind in their HTs, although it is getting easier and more common. It is important to note that while many installers do perform such tests, I have almost never seen one actually do a spatial measurement and average the results from several points (in dB!) when performing a sound system setup. Equalizing a sound system incorrectly can easily result in a poorer frequency response for the system than if no equalization had been performed. I recommend not equalizing a system in the statistical region unless it is going to be done correctly—using spatially averaged measurements and a knowledgeable practitioner. (However I will point out later that it is not the room that we really want to equalize anyway.)

5.2.d Low Frequencies and Damping

It is now time to consider the frequency range below f_s. To do this Figure 5-4 shows a new frequency response curve of the same room, but with about twice the absorption placed on the walls. This damping is ideal-

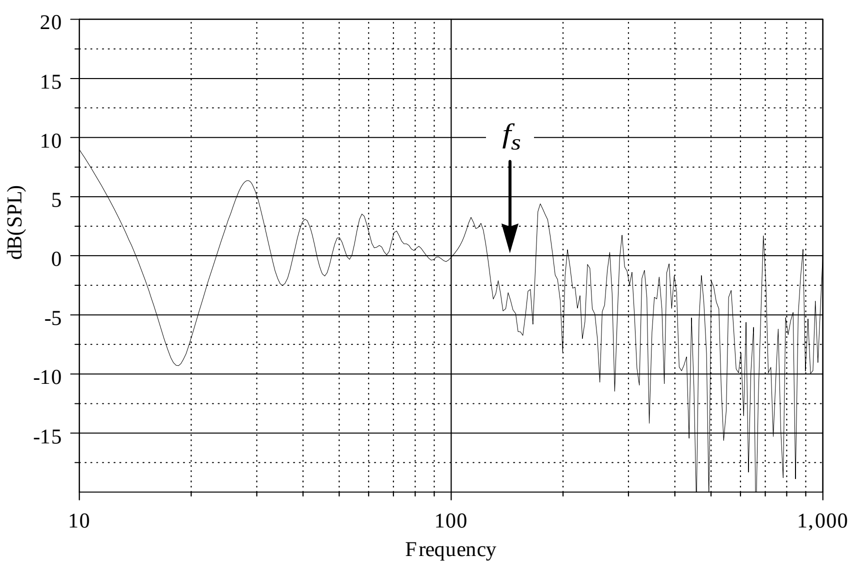


Figure 5-4. Frequency response in a room with high damping—in dB(SPL).

ized in that it is frequency independent, which is far from typical for the types of sound absorption in common use, and would be somewhat difficult to actually achieve in practice. Fortunately, as I will show, this type of frequency characteristic is not really what we will want anyway. It is used here for illustrative purposes only. There are several noteworthy features in this revised figure (compare to Figure 5-4 on page 87). First, the damping has a profound effect on smoothing out the lower frequency ($<f_s$) fluctuations, but it has a notably insignificant effect on the frequency response fluctuations above f_s. Above f_s, the mean level has been lowered, while below f_s there is no obvious reduction in the mean level at all. In the vicinity of f_s, a combination of effects is occurring which can be quite complicated to describe. Clearly, damping does a good job of smoothing out the response curve at low frequencies, but at higher frequencies all it does is lower the energy level.

Figure 5-5 (on pg. 88) shows the one-third octave response for the same room as shown in the previous figure as a direct comparison of the lightly damped, medium damped and heavily damped situations. These curves are all spatial averages of four locations. There are some important features to note in this figure. The lower energy level at the higher frequencies is quite apparent in the figures.

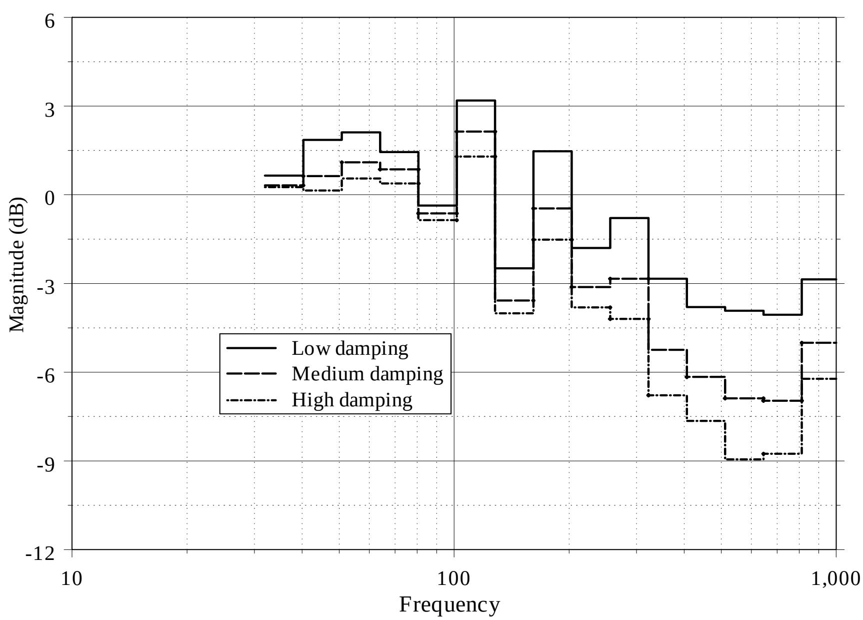


Figure 5-5. Room response with increasing amounts of damping

At the lower frequencies, the increased absorption appears to affect the total energy level by only a small amount, 1–2 dB. Some smoothing is evident in the one-third octave plot, but the smoothing of the response fluctuations at these lower frequencies is much more evident in the narrow band curve shown in Figure 5-4. The damping has improved the 100–200Hz problem somewhat, but this region is still in need of some further attention.

The significant point here is that broad spectrum damping is mostly positive at the low frequencies—up to the point at which the statistical region begins to dominate—but at the higher frequencies no apparent benefit is evident from damping. There may be other reasons to use damping in a room, but from a steady state frequency response standpoint, the only benefits appear to be at low frequencies.

One other thing that I would like to point out before leaving the discussion of these figures is the strong inclination that one might have to call the peak in the response at 180 Hz and just below 700 Hz in both the narrow band room response curves (Figure 5-1 and Figure 5-4) a “mode” or a resonance—it is neither. It is simply a random event which has occurred in the frequency response curve at a particular frequency (unlikely to happen, but it is always possible—like tossing six heads in a row). Another indication of this fact is that adding damping does almost nothing to this sharp peak, a clear indication that it is not a resonance. There is also the fact that these peaks span several modes to tens of modes, not just one. In a spatially and frequency averaged curve, the lower peak is apparent, but the higher one is not. In fact, as the temperature in the room changes these peaks could actually disappear and reappear somewhere else.

It is important to understand the statistical nature of a sound field in a room. I have found from experience that this aspect of room acoustics is widely misunderstood. Explaining some of the implications of these principle is the reason for so many different curves and so many different ways of looking at the same room. More than anything else that I say in this text, the absorption characteristics that I recommend are the most controversial. This is because they are completely contrary to intuition, and quite contrary to what is typically done in large rooms (and most small ones) and what the marketplace recommends. I will show why intuition and convention in the design and treatment of a small room may lead to a sub-optimal listening environment.

5.3 The Direct and Reverberant Fields

A concept that also needs to be grasped is that of the direct field and the reverberant field. As a sensor is moved further and further away from a source in a free field its pressure response falls at –6dB/DD (Double Distance). We have already used this fact on numerous occasions. This rule only holds true in a room to a certain distance from the source. Eventually the reverberation level of the room takes over and the steady-state pressure response in the room remains constant for greater distances. The transition is gradual. The –6dB/DD field is called the **direct field** and the constant field is called the **reverberant field**.

Direct field
that portion of the sound field that is dominate by the source level—no significant reverberation content.

Reverberant field
the region where the sound level is mostly reverberant energy.

I have shown a plot of this pressure response characteristic in Figure 5-6 for several qualitative descriptors of sound absorption. A rule of thumb that I use is represented by the straight dotted line. The point of intersection of this line with the pressure curves gives the point at which the direct field and reverb field energy level are about equal. Points left of this intersection indicate a dominant direct field and those to the right a dominant reverberant field. As a rule of thumb, the listener should not be in a dominate direct field location because at these locations there will not be a good mix of direct and reverberant energy required for spaciousness (see pg. 91). Note, that for lively rooms, this distance is at a reasonable seating distance from the sources, but for a dead room, one has to be so far away from the transducers as to be impractical or impossible.

The direct and reverberant fields are important concepts that will come up again in the context of noise control. It must be recognized that Figure 5-6 applies only to steady state sound. I should also point out that the statistical character of the sound field that I talked about in the previous section only applies to the reverberant field, it does not apply to the direct field. In a real room, there is always a direct field content until such time as the reverberant field has had sufficient time to build to steady state. The direct to reverb level and the time to reach this mixture are both critical aspects of the perceptual discussions that I will get into next.

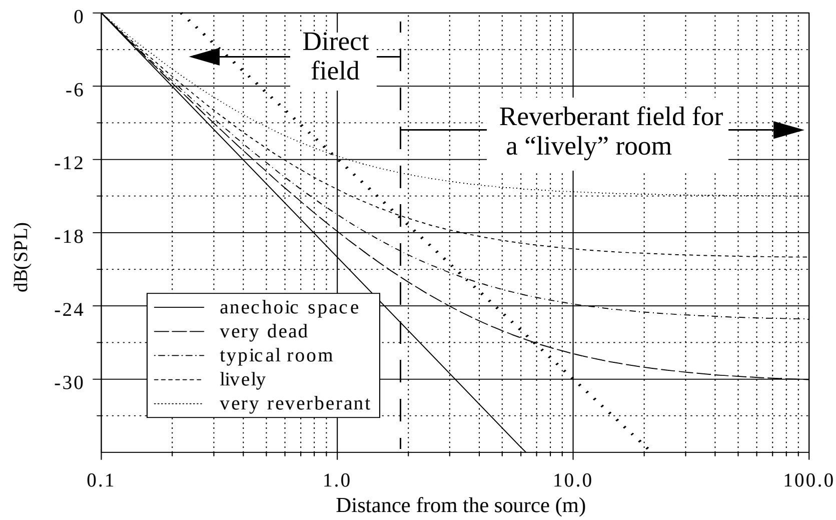


Figure 5-6. SPL versus distance from the source in a room, normalized to 10 cm.

5.4 Sound Perception in Rooms

Before I return to room acoustics design I need to define some requirements from the perceptual domain—drawing on some of the aspects of hearing that I touched on previously. To a large extent, the sound quality of a sound system in a room is characterized by three subjective aspects. They are its spaciousness, its localization capabilities and its timbre or coloration. The room can have a major, almost dominant, influence on these aspects of the sound system and expecting high quality playback without due attention to the room is naive. I will discuss these three criteria for a sound system by first defining what they are and how they are influenced by the room. Then, I will return to the acoustic problems that influence these criteria and address the room design to optimize them.

Spatiousness
the subjective feeling of being engulfed in sound.

5.4.a Spaciousness

Spaciousness is a purely subjective term that applies only to the room and not to the loudspeaker system. Spaciousness is the subjective feeling of being in an acoustically large space with a well distributed sound character.

Precise definitions are not possible, but defining a room with a complete lack of spaciousness is—namely an anechoic chamber (see pg. 81). Spaciousness is created by a large number of laterally arriving sound waves which are delayed from the direct sound by more than 10-20ms. Only the reverberation field can possess this characteristic. The requirement for laterally arriving sound is apparent by considering the lateral placement of our ears and the enhanced ability of resolving sound arrivals in the lateral plan over other planes. Sound arrivals in the vertical plane are difficult for us to determine where they are coming from and, as such, tend to yield a more confused sound perception and are not considered significant or desirable to sound spatiousness.

In order to have the feeling of spaciousness, one must first be in a room location with a reasonably high reverberation level relative to the direct sound level. I showed this point in the previous section and how it varied with room absorption. As I showed in that section, in order to be at a reasonable seating location for good spaciousness, the room must not be heavily damped. A highly damped room causes an increased seating distance and an increased loudspeaker output requirement, resulting from the increased distance and the increased absorption. Spaciousness is just not possible in a heavily damped room. Once the correct distance and room absorption have been located, the time to build up the reverb field has to be considered. This time is both room and loudspeaker dependent. The more directional the loudspeaker the slower the reverberation field will build-up. Actually, calculating all of these requirements are complicated and so it is easier to deal with them in a more qualitative manner.

Loudspeaker Q
the ratio of axial response to the average off axis response.

An omni-directional loudspeaker sends sound in all directions, and, thus, there are many more early reflections from this kind of source than for the directional source. The directional sources takes a far longer time to build up the reverberation field. The directional speaker has a much higher ratio of axial response to power response than the omni-directional speaker. This ratio is sometimes called the **Q** of the polar response (the exact definition is somewhat more complex than this). An omni-directional loudspeaker has the lowest possible Q, namely one. The higher the Q, the more directional the loudspeaker and the slower the reverb field grows in time. This will be seen to be a major factor in the sound system design.

5.4.b Localization

Localization
the ability to sense the location of the virtual sound sources in a recording.

Localization refers to our ability to correctly locate the source of a sound in space. In our context here, it is our ability to visualize, in a hearing sense, the intended or virtual location of the sound emanating from the loudspeakers. This definition is predicated on the idea that the source material actually has a virtual source with an intended precise location. One cannot localize on a reproduced sound if the localization cues are not actually present in the source material. But many recordings do have intended virtual source locations, and the ability of the loudspeaker room combination to accurately place these sources is an important attribute of a quality loudspeaker—and a good room.

Localization is strongly affected by the loudspeaker-room interface situation. Localization in a large room is almost a trivial exercise since there are no early reflections to confound our localization capabilities. In this case the localization capabilities are simply those of the loudspeaker itself. Localization in an anechoic chamber would also be very good for the same reason - there are no very early refection to inhibit the ears localization function. It is easy to see that my three subjective criteria may be conflicting. In other words, good localization may come with degraded spaciousness and vice versa. A compromise has to be reached.

The principle problem with localization in small rooms are the early lateral reflections off of the nearest walls. These refections will be integrated (fused) with the direct sound and tend to confound the perception of the virtual sources location as well as cause coloration of the sound. Vertical reflections do not have a great deal of influence on localization since our localization capabilities are already poor in these directions, but they can have a strong impact on coloration as we will see in the next section.

5.4.c Timbre

Timbre
the tonal character of a sound.

Timbre refers to the tonal shift or coloration of a signal caused by the loudspeaker and room interface. This coloration can occur because of the early reflections, which create a comb filter effect. A comb filter results whenever two signals are added together after one of them has been delayed in time. It is sometimes called multi-path. An example of a comb filter is shown in Figure 5-7 on page 94, where the second signal is delayed from the first by about 1 ms.

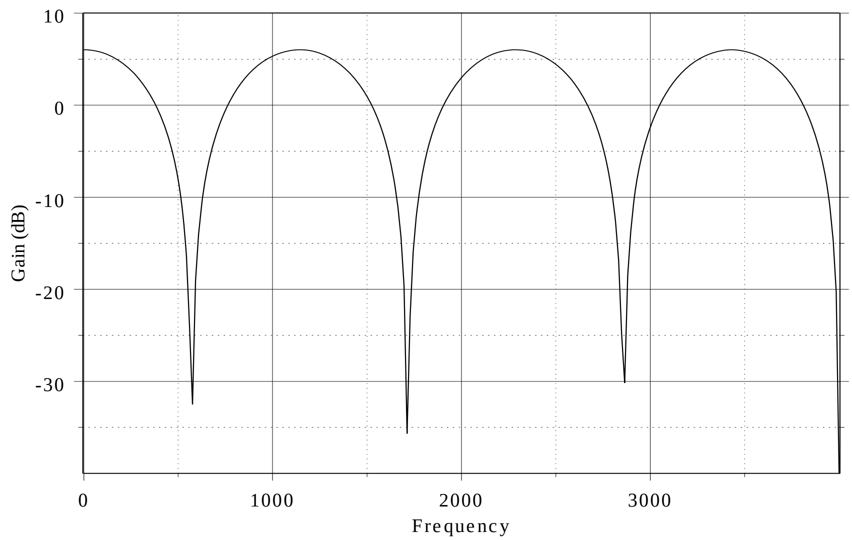


Figure 5-7.
Comb filter resulting from the combination of two signals delayed from each other.

There are two critical aspects of a comb filter for small time delays. The first is that the notches are spaced wide enough that they fall into alternate critical bands, thereby being perceivable as frequency effects. The second is that the small time delays puts the second signal within the ears integration time, thus fusing the two signals together. The ear is actually receiving the filtered response as shown in the figure. When the time delay is extended, the notches become so dense that they are basically not resolvable by the ear, and, second, they are not as readily fused into the perception as the shorter delays. The net result is that the ear is highly sensitive to short time delays, and this sensitivity drops with longer delays.

Timbre is much like localization in that it is strongly affected by the early refections. In one sense, timbre and localization degradation due to time delayed signals are the same thing, but they can also be different. A vertical reflection causes strong coloration but is not a significant localization problem. Lateral refections cause both. Timbre problems can also come from the loudspeaker itself in the form of non-flat power response (where the direct sound and the reverberant sound do not have the same timbre) or strong resonances, which don't tend to affect the localization as much, but do cause significant coloration problems.

It should be pretty clear by now what we want and what we don't want in the way of small room acoustics. Below is a summary:

- First, we want a large amount of sound absorption at low frequencies to help smooth out frequency and spatial response irregularities. This absorption helps what is called the modal overlap or the extent to which the modes interact. Absorption helps to “mix up” the low frequency modes.
- We want as little absorption at higher frequencies as we can reasonably get away with. That’s because, in general, we do not see any positive effect on the frequency response from generic absorption, and we know that it will degrade the perceived spaciousness of the room by removing the desirable multiple lateral reflections.
- We want to try and eliminate early reflections or at least minimize them as much as possible. I have shown how the smaller the reflection delay, the more detrimental it tends to be and extending the reflection delay time while reducing its level is highly desirable. I have not yet discussed how to achieve this, but there are options.
- We will want to be careful not to go overboard reducing the refections because that will simply lead to a dead and lifeless room.
- An early reflection arriving at an alternate ear is not as bad as an early refection arriving at the same ear. The former case has a binaural advantage in the brain’s signal processing that the later case does not. This fact makes most reflections in the vertical plane undesirable, but again we have to balance this requirement against the desire for little high frequency absorption.

Lets now return to our discussion of the absorption in a small room. I have shown that it is desirable to have large low frequency absorption with little high frequency absorption, where there may be a few exceptions used to control specific early refections. In a practical sense, there is a real problem with this requirement. Virtually all acoustical treatments for rooms have large high frequency absorption dropping to almost nothing at low frequencies, which is exactly the opposite of what we want. Clearly, dealing with the absorption aspects of a room by the use of standard materials is not recommended. The use of sound absorption in a small room must be dealt with extremely carefully. It has been my experience that it is almost impossible to make a small room too live at high frequencies. Most typical room construction materials and furniture have significant levels of absorption at high frequencies. Obtaining the right amount of absorption across the frequency band requires different construction techniques and room interior treatments.

I will return to the construction details in a later chapter, however, there are some specific topics that are more relevant here. How absorption actually works is an important issue in our current discussion. There are two principle mechanisms for sound absorption.

The first is to use a porous material such that the sound wave can penetrate it, and, in doing so, the air moving in and out of this porous medium dissipates energy through friction. This mechanism is by far the most common, and there are some specific features to this kind of absorption. First, it becomes increasingly less effective as the wavelength of sound exceeds the thickness of the material. Thin materials will have no low frequency absorption. The second is that since the porous material works on the acoustic particle velocity, the effectiveness of the material is reduced when it is placed at locations of low particle velocity—places like walls where the velocity must go to zero. A piece of sound absorbing material placed on a wall is 1) not very effective and 2) increases in effectiveness as the frequency increases. <u>This most common of all sound treatments is exactly the wrong thing to do.</u>

The second major source of sound absorption is through the actual motion of the room structure—the walls themselves. Of course, if these walls are perfectly rigid—like poured concrete—then this mode of absorption is negligible. But, for a common frame and dry wall construction, wall motion can be quite substantial. Since the wall has mass, its motion will continue to fall as the frequency goes up—that is, unless it has a resilient support structure. All walls must be supported in some way. When the support is resilient (and all supports are to a certain extent, except for maybe a concrete backing), then there will be a resonance frequency and the motion of the wall will fall both above and below this resonance. A wall would typically resonate somewhere below 100 Hz—depending on drywall thickness and the method of mounting. When the wall does move, it dissipates energy through friction. (All absorption is friction of some sort.) The main difference with this type of absorption is that it decreases with frequency rather than increase as the porous material method does. This would seem to be the ideal mode of absorption for a small room and indeed it is. In fact, if done properly tremendous absorption can be achieved at low frequencies with almost no high frequency absorption.

Another concept in sound absorption that comes into play in most HTs that I have done has to do with sound absorption on opposing walls. In my book *Audio Transducers*, I show how, at low frequencies, sound absorption works the same whether it is on one wall of an opposing pair or it is on both of them. By this I mean that the sound absorption is the same whether it is split between two opposing walls or all of it is placed on one wall. This is a good thing to know because it means that if we need to add low frequency absorption to a room, we need only do it on one wall of each of the three opposing pairs. I will show how this is a major advantage when locating a HT in a home.

In small rooms, the obvious preferred mode of sound absorption is to have the walls constructed in such a way that they deliberately move and absorb sound. This technique would never be used in an large auditorium because in those venues we are looking for primarily higher frequency absorption and hanging type materials, such as curtains or drapes are a good choice. In a small HT, the walls should be bare and hard but mounted so that they flex at low frequencies.

5.5 Summary

A complex chapter like this deserves a summary of its main points.

- The first is that the sound field in a room is a random quantity and one must do some form of averaging of samples (frequency responses) to get a valid estimate of the true frequency response.
- The modal region of a room, where there are true resonances, is limited to a relatively low frequency region never more than a few hundred Hertz.
- Damping in a room is effective at smoothing out the modal region but is detrimental to the steady state response at higher frequencies.
- Sound quality is strongly affected by early refections and the level of the reverberant field, and, in a small room, these two requirements can be in conflict.

- The loudspeakers directivity will strongly interact with early refections and the onset of the reverberant field.
- Most commercial sound absorbing materials do not have the correct frequency characteristics for a small room—they work at high frequencies but not low frequencies.
- Wall internal damping is the ideal way to dampen a small room because it can have good low frequency absorption with little high frequency absorption.

CHAPTER 6 NOISE CONTROL

In this chapter, I will give a brief, almost over-simplified, discussion of the topic of noise control. As a general study of noise control, this discussion is grossly simplified, but this text is not about general noise control, but HTs. In general terms, noise control is a massive subject, but, fortunately, there are relatively few topics that we will need to cover that are relevant to HT.

6.1 Noise Measures

Weighting scale
a frequency dependent emphasis place on a sound spectrum.

Noise is measured using the same equipment that is used to measure sound systems. It is also measured using the dB scale and is mostly reported as a dB Sound Pressure Level—dBSPL. However, in noise control work, it is quite common to use a **weighting scale**. The use of weighting scales for noise measures is quite controversial. The current weighting schemes are derived from the equal level contours shown in Figure 2-2 on pg. 34. The problem comes from the same sort of argument that I used when discussing distortion perception. Equal loudness does not necessarily mean equal annoyance. Using a weighting scale for some measures of noise is simply not justified.

Such is actually the case for our situation. If the noise that we are talking about is environmental noise, then annoyance would follow simply from perception. In other words, if I can perceive extraneous environmental noise in my HT, then it will probably annoy me since I don't want any extraneous sounds at all. If the SPL levels in the room from the signal source (the audio system) are measured with a flat weighting and the noise levels are measured with a weighted scale, then totally erroneous results will be obtained for annoyance.

HVAC
Heating Ventilation and AIR Conditioning, the air quality and temperature control system.

One-third octave
a type of measurement filter bank which approximates the resolution of the ear.

In Figure 6-1, I have shown the three most commonly used weighting scales for noise measurements. A-weighting is by far the most common. I will show data from my own home since it is pretty typical. It has a forced air **Heating, Ventilation and Air Conditioning (HVAC)** system and as such has a considerable amount of fan noise. Figure 6-2 shows the narrow band (FFT) data along with the A-weighted version. Note that there is a large noise spike at 120 Hz. This would not be uncommon in a system like this. My HVAC system has already had some noise reduction done on it, so typical results can in fact be much worse than those shown here. As a comparison, a **one-third octave** measurement is also shown in the bottom of the figure. I have shown both these figures to demonstrate several features that one will commonly see in plots of noise. In Figure 6-2, the frequency axis is logarithmic and corresponds to how we would perceive the frequency

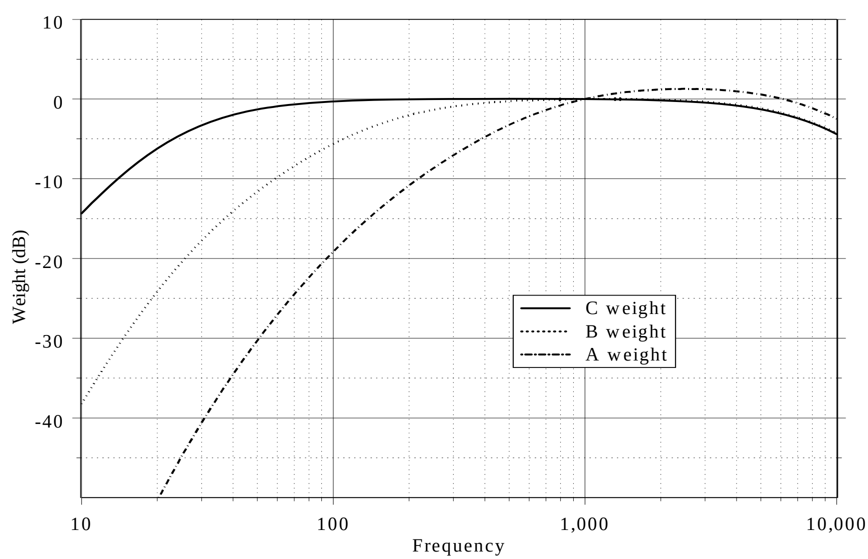
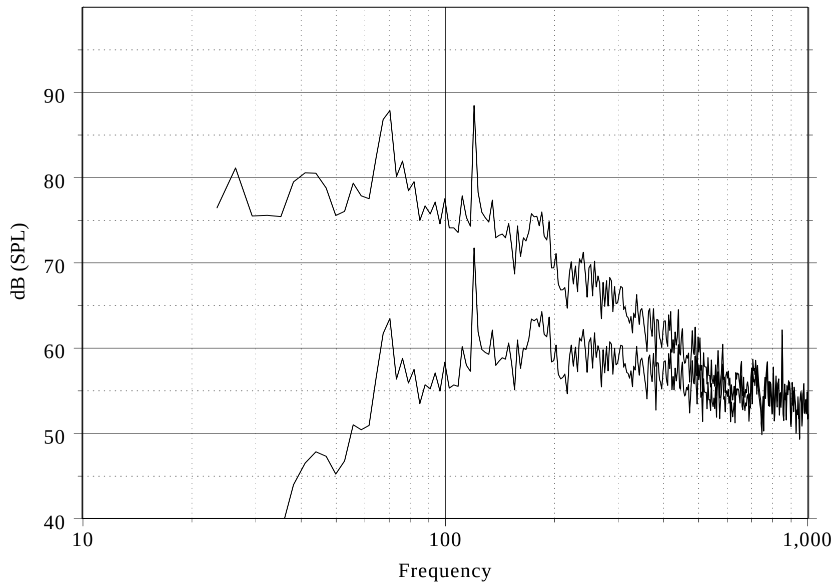
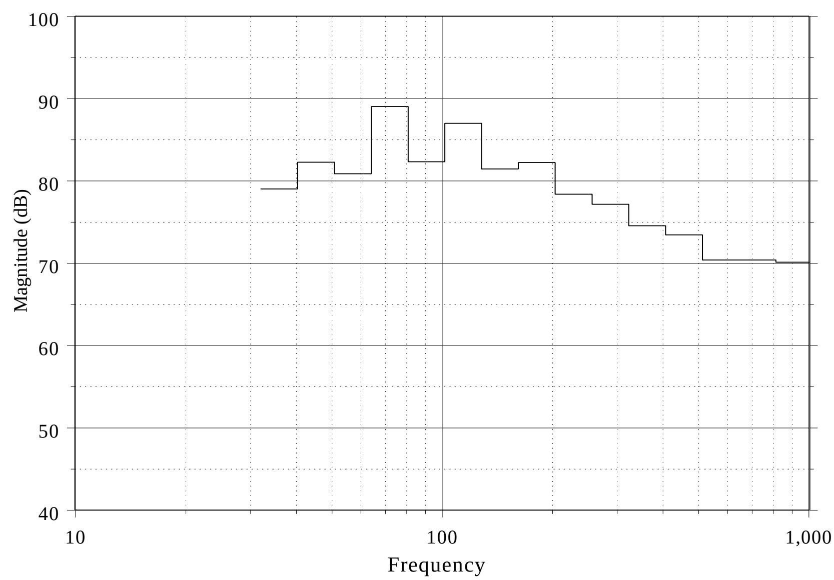


Figure 6-1. Standard frequency weighting scales.

Figure 6-2. Uncontrolled noise level at HT location weighted and unweighted.





scale. Since the FFT is linear in frequency when plotted on a logarithmic axis, there are more points at the higher frequencies than the lower ones. This gives a false impression in the sense of resolution, since you should recall from Chapter 2 that the resolution of the ear is basically constant on a logarithmic scale. That is why the one-third octave plots are useful, the one-third octave bins have about the same resolution as the human ear.

Another characteristic of one-third octave plots is also exhibited in these figures. Note that the two problem peaks at 60 Hz and 120 Hz, which are so clearly evident in the narrow band plots, are shown in the one-third octave plots at two different levels. Even though these two peaks have the same narrow band dB SPL levels, the one-third octave plot show them at different levels. This is a result of the widening bandwidth of these filters, which also accounts for the increasing apparent level of the higher frequencies that results from the greater bandwidth of the one-third octave filters. For these reasons, it is always a good idea to look at both the narrow band and the one-third octave plots and to keep in mind that neither of these plots is exactly the way the ear would perceive the sound.

I want to return now to the problem peaks that I highlighted above. These peaks, at about 88 dB SPL, would be a big problem in a HT. Consider the following scenario. A two way speaker, from Section 4.1.d on pg. 68, having a Max-dB SPL level of 85 dB would not even be capable of exceeding the noise level in this room at these two low frequencies. This is basically a 0 dB dynamic range.

Ambient noise
noise that is always present in a room, but can also include temporary but common noises.

In a commercial theater, the room is usually designed to a very high standard of background noise and achieving these levels would actually be prohibitively difficult in a HT, so, I will take a different approach. Say that we want a 40 dB dynamic range in our room, which is actually pretty bad, but it is what I would consider a minimum, and that we want a Max-dB SPL of 110 dB SPL. That means that I would need a powerful set of loudspeakers <u>and</u> a noise level that does not exceed 70 dB SPL, which is actually quite high, but acceptable for some situations. The **ambient noise** will be perceivable in the quieter passages of a film, but most of the time the movie soundtrack will mask out this background noise. That means that I will need to achieve 18 (let’s say 20) dB of noise reduction. This is a achievable, but it is not at all an easy task. If we wanted a truly quiet room, something more like 40 dB of reduction is required (that’s still higher than a good commercial theater). Since this later amount is actually an enormous reduction—100 times reduction, or removal of 99% of the noise—I usually just recommend going to the max on noise control and not bothering with actual measurements. If you end up with a quieter room than you expected, then it will be a pleasant surprise. If it ends up being too noisy and you did your best, then what else can you do? Move?

I am not, of course, familiar with all home situations and HVAC methods, but I have lived in and been to a great many homes. Almost without exception, every one of them would require an extensive noise control treatment to achieve a truly quiet HT room. Other major sources of noise are home appliances, plumbing (water flow noise—a very difficult problem) and external transportation noise (cars, trains, airplanes). Again, I cannot conceive of a HT that relied on open air ventilation like a window as ever being acceptable. I know of few home locations where this method of HVAC would not be a noise problem. Even the wonderful sound of waves on a beach becomes noise when you are trying to hear a dying man whisper in a film.

The type of HVAC is a significant factor. I had a home with baseboard hot water heating that was actually very quiet. But, to cool this house in the summer required a through-the-wall AC system which was outrageously noisy. The HVAC was fine in the winter, but a real problem in the heat of the summer.

The best thing to do when evaluating your noise situation is to put yourself in the exact seating location and listen carefully. Be sure to do this at various times of the day and under as many different environmental conditions as possible—if the HVAC system operates intermittently, then turn it on. Honestly assess the situation, being careful to recognize that noise not in the presence of a signal tends to be perceived as less of a problem than it is when a desirable signal is present. Again, our hearing mechanism explains this very well—masking of the desired signal by the noise makes the noise more offensive. Consider my previous example. The peaks at 60 Hz and 120 Hz would have a considerable masking effect on the sounds in this room. For example, the masking from these two tones (an estimate from Figure 2-4 on pg. 37) would create a masked threshold of about 60 db at 200 Hz. Even though the noise peaks might only be considered barely audible at this sound level in the room, if any signals at around 200 Hz fall below 60 dB SPL, they will be completely masked. In this context, these noise peaks can be seen as a much more significant problem with signals present than without them. Considering the masking effect is really a much better way of looking at the signal to noise in a real room (as it is with distortion and most things having to do with perception). For the loudspeakers that we have been considering, with a Max dB SPL = 85 dB, we would only have 20 dB of signal to noise.

With noise treatment bringing the levels down by 20 dB, the room will have a masking threshold of about 20 dB (SPL) which is fine, but don't think for a minute that the fan noise will not be clearly audible at this level.

This discussion should be compelling enough to convince you of the almost inevitable need to reduce the existing environmental noise outside of the HT from getting into it. To design an optimal solution to this problem is difficult and I recommend doing as many of the measures that I will describe as possible.

One more thing relative to the statement that I made in Chapter 1 about "one man's signal is another man's noise". The movie track is a signal to those inside the theater, but it can be very much of an annoyance to anyone outside of the theater. Realistic explosions at 100+ dB will send residents not watching the film right out of their seats (and probably into your face). If you want to watch a war movie while your wife prefers to read a book, unless the sound transmission out of the room is reduced by 50 dB or more, she will need to do so in another house!

From this discussion you can see that it is not just that we want to keep sound out of a HT. We also have an equally stringent requirement to keep the sound in. These two requirements combined together are the reason why I say that there is simply no way that one can isolate a HT from the external environment too much. I have yet to see a room in a home that even comes close to being perfectly isolated. Trust me on this—noise isolation cannot be taken too seriously!

My approach to this massive problem is to give the reader an idea of the array of control measures that are possible, what works and what doesn't. Detailed descriptions of the actual construction techniques will have to be delayed until Section 3 when I will talk about the design and construction of the room itself.

6.2 Transmission Loss

Transmission loss
the dB SPL reduction of sound from one space to another space.

Port
Webster's definition — an opening in an enclosure.

Consider a room made of concrete sitting in hard ground with no doors or windows—the ideal isolated room. Making this room livable creates a myriad of short circuits to this ideal. The measure of how much sound isolation that exists is called the **Transmission Loss (TL)**, which is measured in dB. TL is defined as the difference in dB levels on one side of a surface to the other. In the case of our concrete enclosure, the TL would likely be 80–100 dB and would be about the same for all of the surfaces. If I were to put an opening into this room (so that I could actually get inside), then, on that wall, the TL would drop to only a few dB. I could put a door in this opening, but the TL of doors vary enormously with the type of construction. In Table 1, I have shown some typical numbers for TL of various **ports** in a room. First, these numbers are estimates from experience (not exact measured data), second, TL varies considerably with frequency and these are mean numbers. TL almost always falls off dramatically at low frequencies, typically to almost nothing. When determining the TL of a combination of ports, it will usually be the lowest number that wins (looses), i.e. the lowest TL short circuits all other measures. If two ports are about equal, then the net result will be 6 dB less than either one. This makes a high TL extremely difficult to achieve.

Table 6-1 **Transmission losses for typical components**

Port	Range of TL (in dB)	Port	Range of TL (in dB)
interior door	6–10	single pane window	12–25
exterior door	30–40	double pane window	20–35
open door	2–3	open window	2–3
typical Gypsum wall	20–30	special wall	30–50

There is a lot of data and design guidance for wall constructions that isolate two abodes, which certainly makes sense. Unfortunately, there is almost no information on TL for walls or ports between interior rooms. Such a thing is almost never considered. As I pointed out before you had better consider the transmission loss in your HT or you'll be sorry.

But wait—there's more! Any room used for human occupation has to have fresh air into and out of the room. These ports have two problems associated with them. In order to move a large volume of air, I need a large duct. If untreated, these ducts have sound transmissions that vary with frequency from almost 0 dB at resonances to maybe 20–30 dB at other frequencies. Although conceptually possible, in practice one does not tune the duct work for maximum attenuation, it is not practical. There are large amounts of sound leakage from inside of the ducts out, as well as the other way around. There is also the sound that is directly radiated into the room from the air moving device. The TL of HVAC ducts can pose a major design task and one that, if ignored, could result in only a minimal total TL and high levels of HVAC noise in the room itself.

Before treatment, one of my HTs created high sound levels in an upstairs bedroom two floors up. It was quite obvious how the noise was getting in—the air register. Fortunately, this was correctable in my case. But after the construction of the HT, this kind of problem cropping up can often be impractical to solve and one has to live with it. That would be a major disappointment in an expensive installation.

For these reasons, I recommend that the HT have as few ports as possible in order to minimize the need to ensure that each of these ports has a sufficient TL. One door, plus the HVAC connections, is all that is required. Everything else is superfluous and a potential problem.

6.2 Sound Absorption Versus Sound Reduction

The typical approach to noise control is to make the room "quiet". This term appears to be synonymous with a well damped room. Unfortunately, the two things are really quite different. Ignoring for now the detrimental effect that sound absorption has on a room from the psycho-acoustical point of view, I will now consider the effectiveness of sound absorption as a noise control mechanism.

Sound, from a noise source, just as with any source (see Section 5.3 on pg. 90), has two fields, the direct field and the reverberant field. Sound absorption will reduce the sound level of the reverberant field, but it does nothing to reduce the level of the direct field. At some point sound absorption simply becomes ineffective as a noise reduction technique. From Figure 5-6 on pg. 91 we can see that in a typical room, at a listening position of about 3m from the noise source, making this room very dead would only reduce the noise level by about 3dB with this technique. Making the room anechoic would only reduce the level by 9 dB. As a rule of thumb, I would suggest that you estimate the noise reduction from room damping, even in a well damped room, to only be about 6dB at most. For many noise problems this is barely scratching the surface.

To get to a truly quiet room, one must reduce the noise from the offending sources before it enters the room, because once it is in the room it is difficult to get rid of. This direct attack on the noise source can also be a difficult proposition as it can involve the modification of some complex home components, mainly the HVAC system. I will now discuss some basic noise control elements and I will show how to use them in later chapters.

6.3 Duct Noise

Noise that propagates in the HVAC duct system can go anywhere and come from anywhere. The best approach to this situation is to modify the entire duct system altogether. Some of these modifications are straight forward and simple while others are much more difficult to implement.

Soft-sided ducts
a flexible tube made of a coil of wire attached to a thin plastic sheet surrounded by fiberglass.

The first thing that I always recommend is to replace sections of every arm of the existing system with a **soft-sided duct**. To be most effective, this should be done to all of the legs, not just ducts into the space of the HT. This material is readily available and not too expensive. It is basically a coiled wire expandable frame wrapped in a thin cellophane type liner which is then wrapped in a blanket of fiberglass followed by a PVC cover. The cellophane layer of this material actually lets the sound fluctuations pass through the walls into the sound absorption material, but of course some of the sound will actually continue to pass through the PVC layer to be reradiated. Normal sheet metal duct work does not absorb any of the sound energy except as sound leakage (re-radiation).

This material does wonders for the HVAC system throughout the entire house since it basically converts the highly reflective duct system into an highly absorptive one. This means that sound will no longer flow between the noise sources and all the other rooms, basically unattenuated, via this transmission system. It is amazing how effective this fix can be both for the HT and the rest of the home in general. The noise level from the HVAC system can be reduced by some 10 – 20 dB by simply replacing sections of duct, which are typically 6'–8' long, with a section of the absorptive duct. It is not necessary to replace all of the duct work. A patch of a several feet in each leg is effective.

Muffler
a device which allows a static air flow which is unimpeded but blocks sound

For the specific ducts going into and out of the HT, another additional component should be added. This component is a **muffler**, completely analogous to the muffler on a car. In order to allow for a large flow of air without a large flow of noise we will need these mufflers in both the inlet and outlet ducts for the room. It is a good idea to keep these inlets and outlets to a minimum in order to minimize the number of mufflers that have to be built. If multiple ports are required, then it is best to make them on a common duct leg that will allow for a single muffler. In the chapter on construction techniques, I will describe how these devices can be readily fit into most existing construction. Here I will describe the general theory behind their use.

The basic idea of a muffler is to use the energy reflection from a cross sectional change in a duct to create an acoustic filter. Mufflers can be extremely complex (as in vehicles), but for our purposes we will only consider the simplest designs as shown in Figure 6-3. The muffler has an inlet

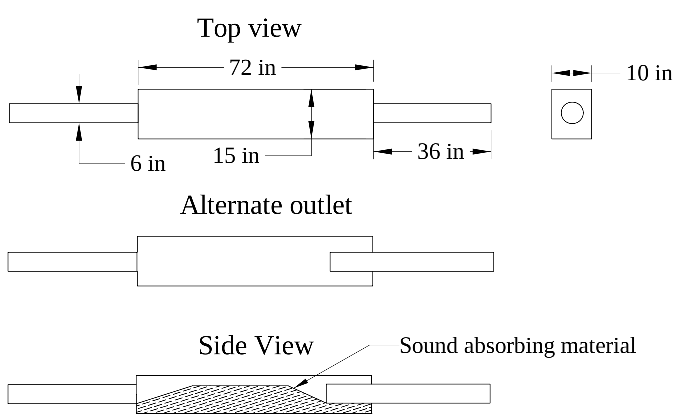


Figure 6-3.
A simple muffler example showing alternate inlet options and location of sound absorption.

and an outlet with areas S_1 and S_2 respectively. Typically these areas are the same. In the first example it doesn't matter which is which. In theory the effectiveness of a muffler is given by the mathematical equation:

$$\alpha_t = \frac{4}{4\cos^2(kl) + \left(\frac{S_2}{S_1} + \frac{S_1}{S_2}\right)^2 \sin^2(kl)}$$

Power transmission the amount of sound power transmitted through a muffler.

Insertion loss the number of dBs of sound attenuation provide by inserting a muffler.

where α_t is called the **Power Transmission**. A more useful quantity is the **Insertion Loss** (IL) which is the actual dB reduction that is to be expected by inserting this muffler. Insertion loss is defined as:

$$IL = 10\log(\alpha_t)$$

For the example muffler shown in the top of Figure 6-3, the IL predicted by this equation is shown in Figure 6-4.

In the variable $k = 2\pi f/c$, c is the speed of sound and f is the frequency. The length of the muffler is denoted by l and the combination kl is a dimensionless number indicating the number of wavelengths in the duct length.

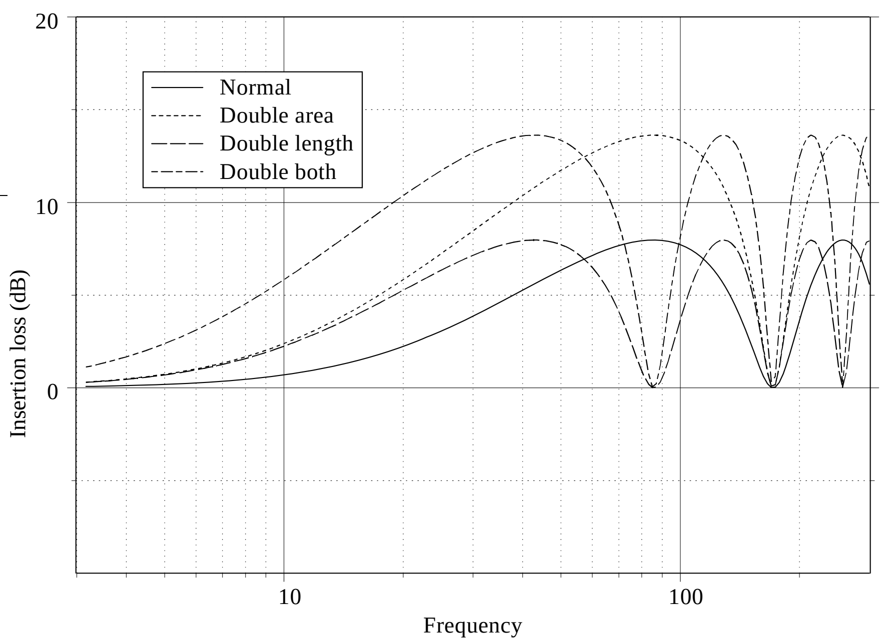


Figure 6-4. Theoretical Insertion Loss for an example muffler.

The muffler has no effect at the lowest frequencies, because it still has to allow for the static airflow. The effectiveness increases with cross sectional area of the muffler and with the length of the muffler—basically, bigger is better. It can be seen that there is always a frequency for which the IL is zero, indicating no loss at all. In complex muffler systems, this effect is usually neutralized by using series mufflers of different lengths.

For a real muffler the situation is not so simple, because a real muffler has an outlet duct of finite length. This finite length causes resonances to occur in the muffler system and some substantial deviations occur from the ideal muffler. These can be noted at the lower frequencies in Figure 6-5. This figure shows the calculated results for the muffler with a finite length of exit duct for three different outlet duct lengths. This figure also has taken into consideration some sound absorption material in the muffler. Note that there is a resonance that actually causes the IL to go negative—an increase in sound transmission. This could be a major problem if located at the wrong frequency. Notice that for longer outlets the damping basically causes the resonance gain to go away. The point here is that the length of the exit duct must not be made too short or problems could ensue. In the

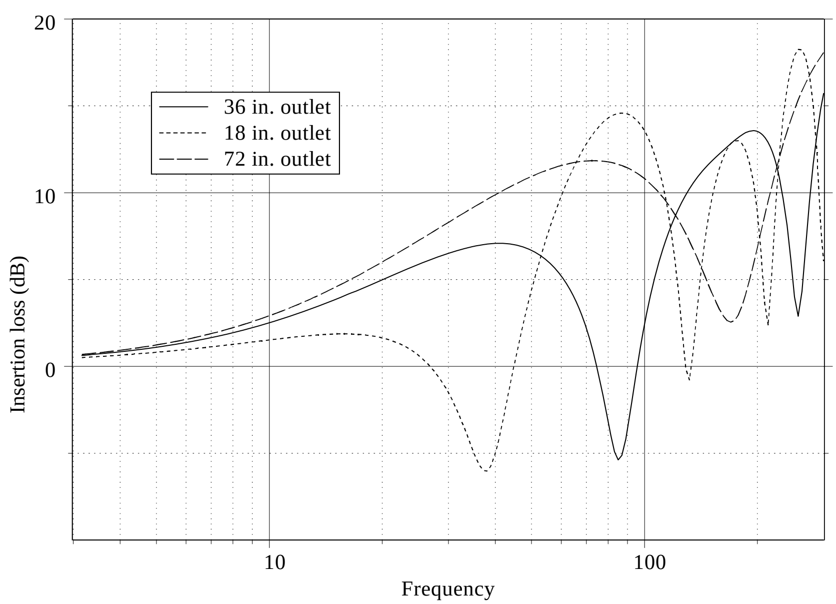


Figure 6-5. Insertion loss for a 12” x 16” muffler on common 6” ducts with a 3' outlet duct (solid). Variations of the duct length are also shown.

second example, the duct has been extended into the muffler body thereby making it longer, but at the sake of a slightly shorter virtual length of the muffler body.

In Figure 6-5 you should also observe that the IL increases quite substantially at the higher frequencies due to the absorption. This is a positive effect and one that will actually continue to increase with increasing frequency. IL values of 30–40 dB can be achieved at higher frequencies in a well damped muffler. I should give a word of warning here. This absorption must not be allowed to interfere with the actual airflow or the effectiveness of the duct at supplying ventilation air could be severely impaired. It takes very little restriction in a typical HVAC system to virtually cut-off the flow. This once happened to me by accident when the absorption material moved after installation. I had to disassemble the unit to correct the problem because the room was otherwise unusable.

From all of these curves it is apparent that the practical low frequency IL limit to this size muffler is about 10 dB on the average. This is not very effective in a frequency range where there is not only a lot of fan noise but often a lot of sound track signal. In every instance of a HT construction that I have noted—the HVAC system is usually the weak link. The HT in my current home can still be heard (at an acceptable level) at the HVAC duct in a bedroom two floors away, even after substantial treatment. Preceding without installing a duct muffler can virtually eliminate the effectiveness of all other noise control measures. In other words, if you have a duct HVAC system and you do not install some form of muffler (or at a minimum soft-sided duct work) on the system, then there may not be much point in doing anything else. Don’t short change this muffler in terms of space or quantity (every inlet and outlet must have one) or the overall effectiveness of the noise control measures can be severely diminished.

6.4 Wall transmission

I want to end this chapter with a discussion of sound transmission through partitions—walls, windows, etc. This topic is, in general, a massively complex subject, but there are a few things that I would like to discuss because they are so important.

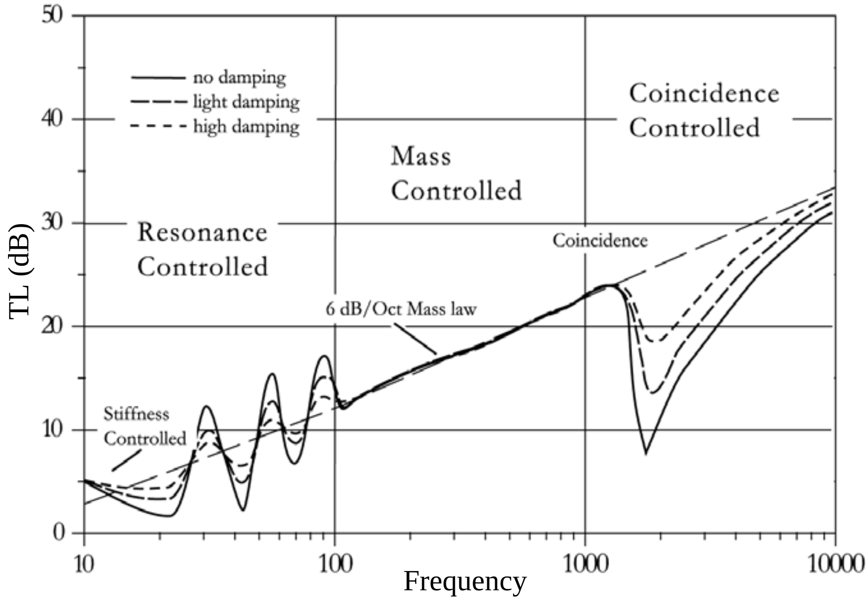


Figure 6-6. Transmission Loss for a typical partition.

Resonance, mass, coincidence controlled three regions of sound transmission characteristics for a panel.

Coincidence where the wavelength of sound in air matches that in a panel.

All partitions behave in a basically consistent manner with three distinct regions of operation. From a Transmission loss viewpoint, a typical partition would look similar to that shown in Figure 6-6. There are three dominant regions, the **resonance controlled** region, the **mass controlled** region and the **coincidence controlled** region.

Coincidence occurs when the wavelength of sound in air matches that in the partition. Since the wave speeds are different in the two mediums this occurs only at a particular frequency which is marked on the graph. The wave speed is constant in air, but it is not in a panel for different angles of sound incidence. This means that coincidence occurs over a range of frequencies, not just one. The partition that I have shown in the figure has a fairly high coincidence frequency which was drawn this way to distinctly highlight the three regions. The stiffer the material of the partition, the higher the coincidence frequency. Note how the transmission is highly dependent on the damping in this region. The damping in this curve is internal damping of the wall and not damping material placed on the wall. The internal damping is not appreciably affected by damping material placed on the wall. Internal damping of the walls has a positive effect, and we will want to make it as high as possible.

The TL for a well damped partition becomes close to following the simple mass law. In other words, once the internal damping is sufficiently high the only factor controlling the transmission loss is the mass. The higher the mass the greater the loss. In the past, sheets of lead have been hung as an effective noise barrier—a high mass loss. Today, that is not an acceptable solution.

The resonance controlled region depends on many factors including material mass, stiffness, and, most importantly, the wall mounting techniques. The stiffer the mounting the higher in frequency the resonance region will go, which is neither good nor bad if the wall is well damped. But, if the partition is undamped, then the partition will have a highly irregular TL. Since loss through the wall is sound absorption to the room, this uneven TL will cause an uneven absorption coefficient in the room. It may not be clear as to why this would happen so let me explain.

In the mass controlled region and above the wall is basically rigid to impinging sound waves. If the surface is acoustically untreated, then it will reflect virtually all of the sound that is incident upon it, i.e. it is a reflector resulting in a high transmission loss. At coincidence there will be some apparent sound absorption, but since the wall is basically not moving the absorption is still pretty low. But, in the resonance controlled region, things are quite different. Since there is little TL in this region (low mass loss), most of the sound impinging upon the walls goes right through it. In real rooms, this is the dominate mode of absorption—the fact that the sound simply leaks out. From Figure 6-6, we can see that in the resonance controlled region the apparent absorption will be much higher at some frequencies than at others due to the resonances. Therefore, it is most important that the wall be as internally damped as possible. This has two desirable effects. First ,is the fact that the energy is absorbed by the internal damping and not re-radiated on the other side as noise, and, second, that the absorption effect is smoothed out—creating a fairly smooth broad band effect. The high internal damping also increases the TL at the higher frequencies.

The method of attachment also has a lot to do with the TL of the wall. At low frequencies, the sound pressure in a room acts on the wall almost uniformly over its surface. This pressure causes a large force to appear on the wall and the wall will move. If this wall is rigidly attached to a flexible structure — like wood studs — then the entire structure will move and propagate sound energy throughout the rest of the house as structural vibrations. This is by far the most difficult mode of sound transmission to control as there are few effective localized solutions (there is one that I will show in the chapter on construction techniques). Reconstructing the house is usually not an option. However, there are a few evolving and most interesting new home construction techniques that the really serious HT practitioner might want to consider. Things like prestressed concrete floors and steel framing would be an enormous improvement in the structure borne noise problem, but these techniques are not yet common nor would they be inexpensive.

The bottom line in this section is that we may have actually found a good solution to our room absorption problem. By using flexible but heavy wall structures with high internal damping the ideal of high low frequency absorption with low high frequency absorption can be achieved. In later chapters, I will show how to construct such walls. There, I will show how this construction is completely different than standard wall construction and that the standard construction is completely wrong for our purposes. In general, typical construction techniques are poorly damped, rigidly connected to the structure, and of relatively low mass.

6.5 Tuned Absorbers

Tuned absorber
a resonant system where the high energy levels at resonance are used to absorb sound.

To end this chapter I will discuss a common technique for low frequency sound absorption that can be effective if done properly. A **tuned absorber** uses a resonant system to create high energy levels in a narrow frequency band which can then be absorbed more efficiently because of the higher vibrational energy level. A typical absorption characteristic is shown in Figure 6-7. This set of curves have been normalized to the resonance frequency at a value of 1.0. The two solid vertical lines represent a one octave width.

Absorption coefficient
a value which indicates the amount of incident sound that is absorbed.

All resonant absorber systems have the characteristic that as the bandwidth of the absorption is widened with more damping, the effectiveness decreases rapidly. An **absorption coefficient** of about .7 (70% of the incidence energy is absorbed), on average, can be achieved only over a narrow

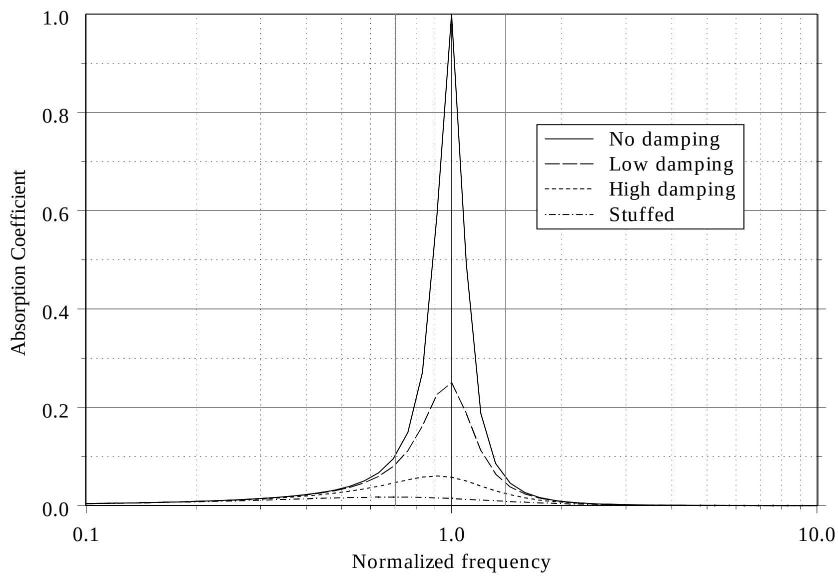


Figure 6-7. Absorption characteristic for a tuned absorber normalized to its resonant frequency.

one-third octave bandwidth. These narrow absorbers can be effective at attacking a pure tone noise problem like that shown in Figure 6-2, but as a room absorption treatment this is far too narrow.

The resonator can take many forms, but the most common would have to be the Helmholtz resonator. Think of a **Helmholtz resonator** as a milk bottle. It has a neck and a volume that form a mass and spring system that resonates. It has a simple tuning equation:

Helmholtz Resonator
a simple air mass and spring system like a milk bottle.

$$f = \frac{c}{2\pi}\sqrt{\frac{S}{l'V}}$$

where c is the sound speed, S the neck area, V the enclosed volume and l', the effective length of the neck. The effective length is always difficult to determine but is usually about 15% longer than the physical length (a rule of thumb).

Panel Absorber
a panel backed by air which resonates and acts as a tuned absorber.

It is also possible to have the mass function of the air in the neck replaced by a panel, which creates what is called a **panel absorber**. The design of this type of absorber can be fairly complex, but these panel absorbers can also be quite effective. The interested reader is referred to the omnipresent web for further study of this subject.

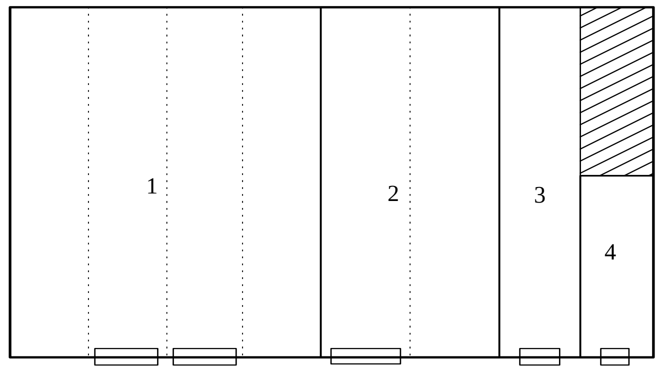


Figure 6-8. Example of Helmholtz resonator construction.

I have used an array of narrow tuned absorbers to cover a wide range by using the volume of space beneath a riser at the back of the HT. A sketch of this is shown in Figure 6-8. Each line represents a joist which did not go all the way to the floor. Where the lines are solid, there is a fixed partition beneath the riser. This created four cavities, each with twice the volume of the previous one. The necks, or ports, were also doubled in area for each cavity. This gave a doubling of tuning for each section. The interior space had 3" of fiberglass insulation placed on the floor of each volume. The combined result of this design is shown in Figure 6-9. Tuned absorbers can be designed into a wide variety of available spaces.

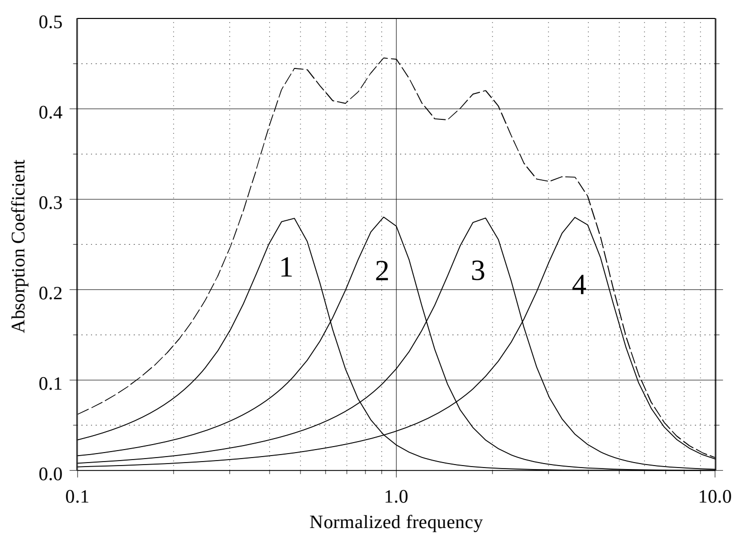


Figure 6-9. Absorption effect of staggered tuning of four Helmhotz resonators. 1.0 equals about 30 Hz in this graph.

CHAPTER 7 IMAGES, COLOR AND VISION

Just as I started out my discussion of audio with a description of the audio receptor, the ear, I will start out the video sections with a discussion of the eye and the basic theory of color. The eye tends to be a bit easier to comprehend than the ear owing to the more linear form of this system.

7.1 Basic Concepts

7.1.a Imaging theory

Physically the eye is so much like a camera—a device that is well know to everyone—that I will begin with a discussion of general lens theory which is applicable to a camera as well as the eye. I will point out a highly important characteristic of cameras and lenses that is not widely appreciated, but is of fundamental importance in the processing of video signals.

Lens
an optical device, usually made of glass, which performs various optical filter functions on an image.

A **lens** is the principle device for imaging in any optical system. It can be thought of as a subsystem with an input and an output where the basic equations are analogous to audio systems but in two dimensions. That's because the input to such a lens is two dimensional, just as its output. The transfer function for a lens has both an impulse response and a spatial frequency response (just like an electrical filter, see pg. 24), but these characteristics must be thought of in two dimensions. I cannot over emphasize the importance of this similarity.

It also turns out that a lens basically performs a two dimensional Fourier Transform (see “Video Signals and Systems” on pg. 24) on an image as the optical signal is transferred through the lens. A 2D image placed on one side of a lens, and closer than the focal length of the lens, will actually produce the Fourier Transform of this image in the plane located at the focal length on the other side of the lens. If the image is moved off to infinity (or nearly so), then this turns out to be equivalent to a second Fourier Transform being performed on the image. This second transform results in an exact reconstruction of the image in the focal plane on the other side of the lens (but with the directions reversed) This is the main feature of a lens with which we are all familiar. For any image to lens distance, it turns out that there is a image plane distance on the other side of the lens, as well as a Fourier Transform plane in which the two dimensional Fourier Transform of the image will exist.

This discussion has been theoretical thus far and not really too relevant or useful, at least not directly. The reader is probably wondering why I have gone into this depth here. I am trying to show how a spatial video image can be equally well represented in both its image plane and its Fourier Transform or frequency domain plane. This fact is completely analogous to the time and frequency domain representations of a sound signal, and, as such, all of the discussions about transforms and domains in the audio section will apply equally as well in the video portions of the book.

It also turns out that a spatial frequency domain representation of a signal (an optical image) is a most useful one and that the quantification of a video system is more efficiently done in the spatial frequency domain. Unfortunately, the optical frequency domain is not as intuitive as it is in audio. Since color is a frequency/wavelength proportional quantity, an optical system actually has two frequency responses. One is its spatial frequency, which I talked about above and the other is the frequency of the light itself which is the actual color of the light wave. These are two completely different frequency domain aspects of the problem, and both are important. One affects the color tone, which is analogous to the timbre of an audio signal, while the other is a spatial characteristic that defines the systems spatial resolving capability which is also completely analogous to the spatial sound radiation problem and the polar response.

As a simple example of the spatial frequency aspect of the optics problem, I have shown in Figure 7-1 four two dimensional images with single spatial frequencies for three different examples (I will discuss the color frequency aspects in the next section). The figure shows a spatial *x*, a spatial *y* and spatial *x* and *y* images. Also shown in this figure is an example of the effect of filtering out spatial frequencies (above $10\,\mathrm{m}^{-1}$ in this example). In this figure I am considering only the intensity of the light not its frequency (color).

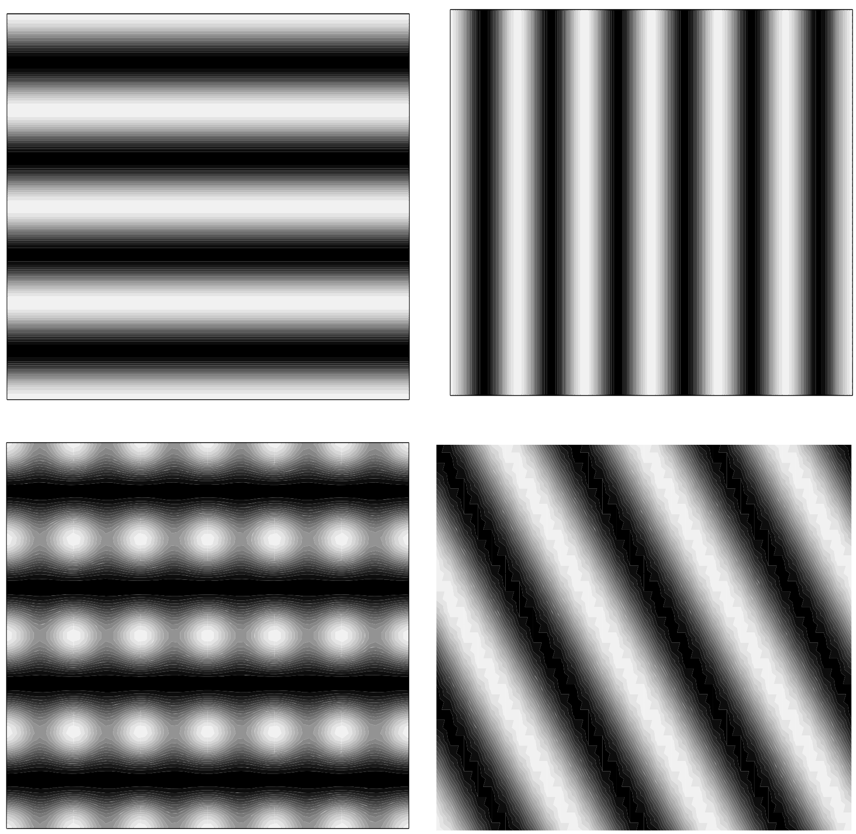


Figure 7-1.
An intensity plot for four different spatial frequencies k. The upper left is a spatial frequency in the y direction of $k_y = 4$*. The upper right has* $k_x = 12$*. The lower left shows two simultaneous waves of* $k_x = 12$ *and* $k_y = 8$*, with the* k_x *wave amplitude at .5 resulting from filtering. The lower right shows a single k where* $k_x = 3$ *and* $k_y = 1.5$*.*

Higher spatial frequencies represent higher rates of spatial change in the intensity of the image. In this sense, the spatial frequency represents the image's detail. An image with predominately low spatial frequency is smooth without much detail, while one with high spatial frequencies has fine detail. An optical systems spatial frequency transfer function then tells us how the spatial frequencies are transferred through the system. Like audio systems, all optical systems are inherently low pass filters which represents the capability of the system to transfer image detail. This characteristic is similar to the better know term **resolution**, although, as we will see, the latter is more aptly applied to a digital optical system.

Resolution
the ability of an optical system to process fine detail—high spatial frequencies.

If an optical system had a cutoff (one-half power point) at a spatial frequency of $10\,m^{-1}$, then the two frequency images above would have a predominance of the horizontal lines in it since the vertical frequency would be substantially lower in amplitude. These plots have been normalized for a constant output level, and it is also not realistic to have different filters along different axes (although this is possible), so the actual plots are somewhat unrealistic. I did this simply to make a point.

Modulation Transfer Function
the optical means for determining the frequency resolution—spatial frequency response—of an optical system.

The actual values that are used to determine the spatial frequency transfer amplitudes is called the **Modulation Transfer Function** or **MTF**. An optical system's resolution capability is shown by plotting the value of MTF versus the spatial frequency, much like an audio systems is its frequency response. In order to understand the meaning of the MTF, consider Figure 7-3.

In this figure, I have shown a horizontally modulated image signal and three "apertures", the size of which represent three levels of spatial resolution. The top is a high resolution, the middle is a medium resolution and the bottom is a low resolution. As these circles are swept across the image horizontally, the average level within the circle is determined. The values thus obtained are shown by the three curves on the right. The top curve has a value of about .9 and the bottom curves a value of about .1. The modulation level of the averaged signal will decrease as the resolution aperture decreases.

Contrast Transfer Function
the MTF for equally spaced solid lines of constant width, light and dark.

In practice, it is easier to define an test image pattern which consists of vertical (or horizontal) equally spaced light and dark regions—lines. In this case, the measurement is actually called the **Contrast Transfer Function**

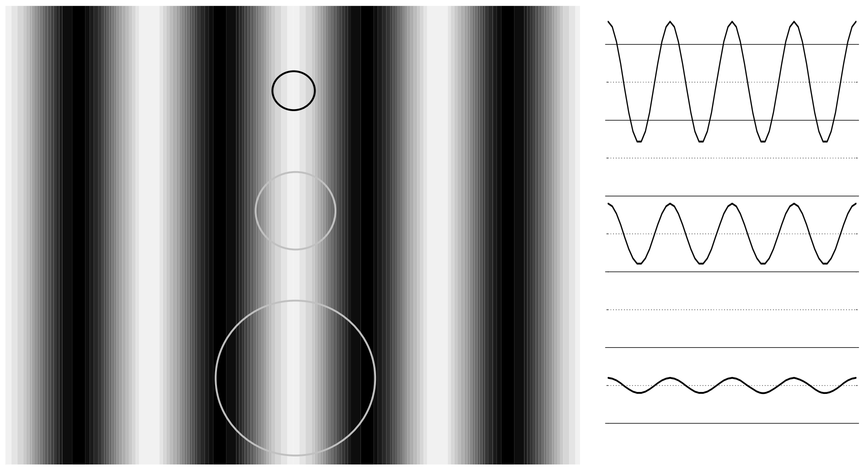


*Figure 7-2.
A graphical representation of the Modulation Transfer Function*

TV Lines
the number of CTF lines that can be resolved in the image.

CTF. MTF and the CTF are closely related to one another, but CTF is the one more commonly used in practice. In fact, in the broadcast industry the standard term for the spatial frequency is **TV Lines**, or the number of resolvable lines in the image. This definition then is independent of the physical dimensions of the image. I will use this latter designation, TV Lines, in this text. TV Lines are basically the same as the pixel number in a computer system, In other words, a screen with 640 x 480 is capable of 640 TV Lines vertical and 480 horizontal. Since the picture aspect ratio is 4:3 the actual spatial frequency is the same in the two directions in term of lines per distance.

Charge Coupled Device
a 2D electronic sensor that emits an electrical charge proportional to the light intensity falling on a pixel.

To give you a feel for these new quantities, I have shown an example of a typical optical system in Figure 7-3 on pg. 122.This is a complete camera imaging system composed of a lens with color filters, a **Charge Coupled Device** (**CCD**) video sensor and, finally, the display monitor. The total response of this system is found by multiplying the responses of each of the components (or equivalently adding the dB losses for each component). In this example, the system is said to have a capability of 650 TV Lines.

I will return to the discussion of image resolution in the next chapter when I talk about Video Systems.

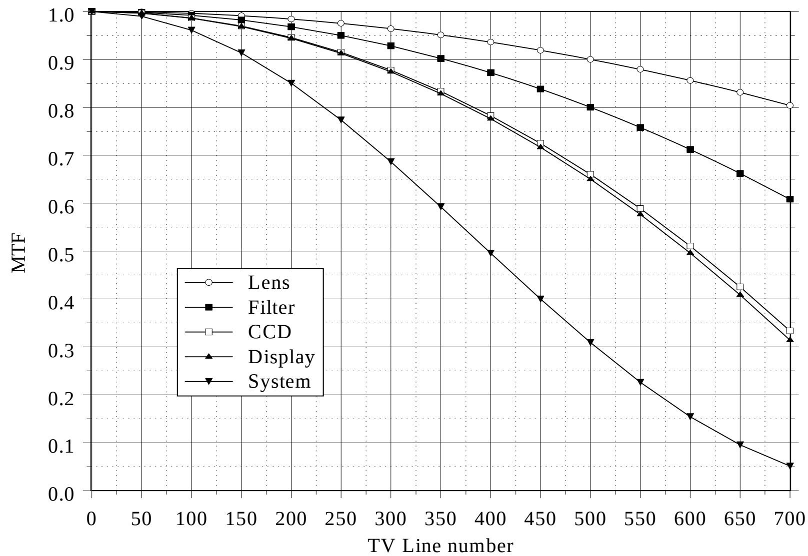


Figure 7-3.
The MTF for various optical subsystems in a typical video system.

As a sideline, I want to describe some fundamentals of good lens design which you should now be in a position to understand. Optical lenses made of glass are rigid and they have a fixed focal length. As such they must be able to move along the lens axis in order to focus on a fixed image plane not at infinity. Unfortunately, glass is also a dispersive medium and it has different wave speeds at different wavelengths (colors), which is the reason that a prism creates a color spectrum of light passing through it. Because of this dispersion, the focal length is color (optical frequency) dependent. This also makes the MTF capabilities of the lens color dependent. In order to get a color independent lens system, multiple elements (individual lenses) are required in order to facilitate a color correction. One uses different lens shapes (convex, concave, etc.) to balance out the dispersive nature of the lens while keeping the average focal length constant.

Therefore, for good color and MTF response, a glass lens must have multiple elements. The larger the diameter of the lens, i.e. the better its light capturing ability, but the more important multiple elements become because the dispersive effects get greater with lens size. Unfortunately, multiple elements come at a cost in light capture and picture quality. Each lens element reflects light at its surface creating multiple internal reflections which not only reduces the light capturing ability (the light transmission) of the lens but also has detrimental effects on its spatial and temporal frequency responses. These multiple element refections can be seen a “lens flare”. The solution is to coat the lens elements with non-reflective coatings to minimize these reflections. The better the lens coatings the more elements that can be used and the better the color correction that can be achieved. Years ago lens coatings were a trade secret and made certain lens companies products clearly superior in performance. Today, coatings are a well know art that all lenses utilize, and they are all designed with a computer optimizer. So much for my sideline into lens theory.

7.2 Color Theory

7.2.a Luminance

When I talk about color theory, I will do so from the technical perspective and not the artistic one. The term color theory is sometimes used by artists, but this theory is completely different than what I will talk about here. As in audio, the term “color” has an ambiguous meaning because it represents both the objective and subjective aspects of the visual image. I will try and relate these two domains together, but this will only be partly accomplished.

Luminance
a measure of the intensity of the light.

Hue
the predominate color in a lights spectrum.

Objectively color can be defined by three attributes, luminance, hue and saturation. In subjective terms, these are brightness, color and purity. **Luminance** is the quantity that I was actually using in the previous section—the intensity of the light from image. For black and white, luminance is the only variable needed to define the image—the 2D image in values of its luminance. To define color, I must introduce two more variables, the hue and saturation.

The **hue** is basically a description of the light spectrums most predominate frequency. The peak in the spectral curve. Obviously, if there is no peak in the curve, then there is no perceived color, i.e. it looks grey. And this brings us to saturation.

Saturation
a measure of how pure the hue is—how much white light versus color light is present.

Saturation is the ratio of the energy in the spectral component to the total spectral energy. This yields a color specification in two variables, hue and saturation, and its overall light output level as luminance. An example spectrum is shown in Figure 7-4. The spectrum can thus be define by its background or "whiteness" level and a major spectral component, if one exists. Obviously, there is no hue for a saturation of 0%, pure white light.

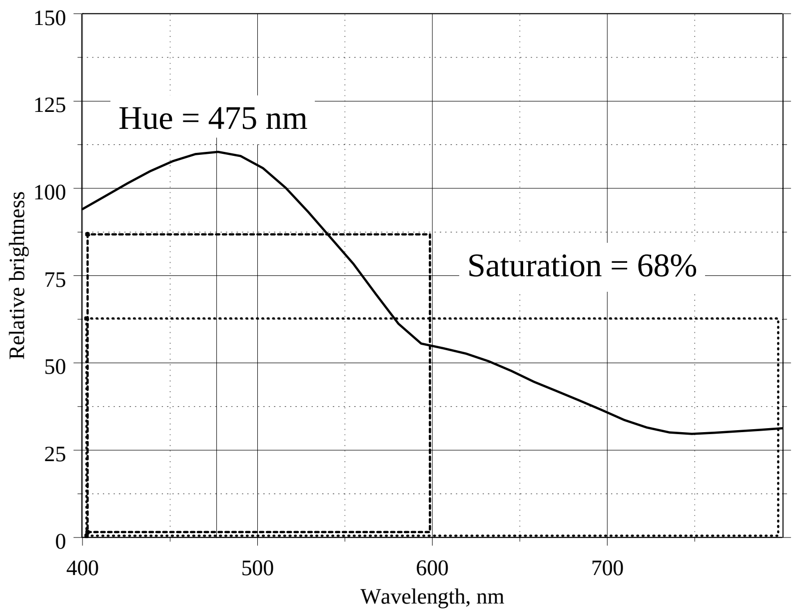


Figure 7-4.
Light spectrum example. The hue is the wavelength of the dominate peak and the saturation is the ratio of the area of the tall dashed box to the area of the short long dotted box. The luminance value would calibrate the "relative" scale in this figure.

7.2.b Three Color Systems

It has always been amazing to me that a color, visual tone, can be uniquely specified by only three variables. No such simplification is possible for an audible spectrum. The reason for this comes from understanding how the eye works.

As in hearing, where we have critical bands of spectral averaging, the eye has a finite amount of spectral resolution. The audio band requires about 32 critical band widths to cover the entire perceptible audio spectrum, but it only takes three bands to cover the visual one. And unlike audible critical bands, the visual ones are fixed in frequency. These characteristics are a direct result of the three cones in the eye—each of which has a spectral distribution.

The three color nature of vision is by now well know and through years of research (from the time of Sir Isaac Newton) exact specifications of these three colors has come about. By performing subjective color matching studies, a set of three curves can be mathematically defined such that any color can be made up of a sum of these three primary colors. This process is called principle components analysis, a well know scientific technique. The functions that result do not actually represent the sensitivity of the three cones in the eye, but they can be thought of as an equivalent set of values which completely defines the subjective space. It should also be noted that these three colors represent what is called the three **primary colors** in an additive model. The other set is the subtractive set used for reflective colors such as in printing, the well know Cyan-Magenta-Yellow (CMY) set. I will not talk about the differences in these two sets since we will be using only the additive model.

Primary Colors
the three colors that make up the color spectrum, there are two sets, additive (RGB) and reflective (CMY).

Figure 7-5 shows the set of functions know as CIE chromaticity coordinates, which were derived by the procedure defined above. These curves represent the colors of the optical filters for a three color additive model of

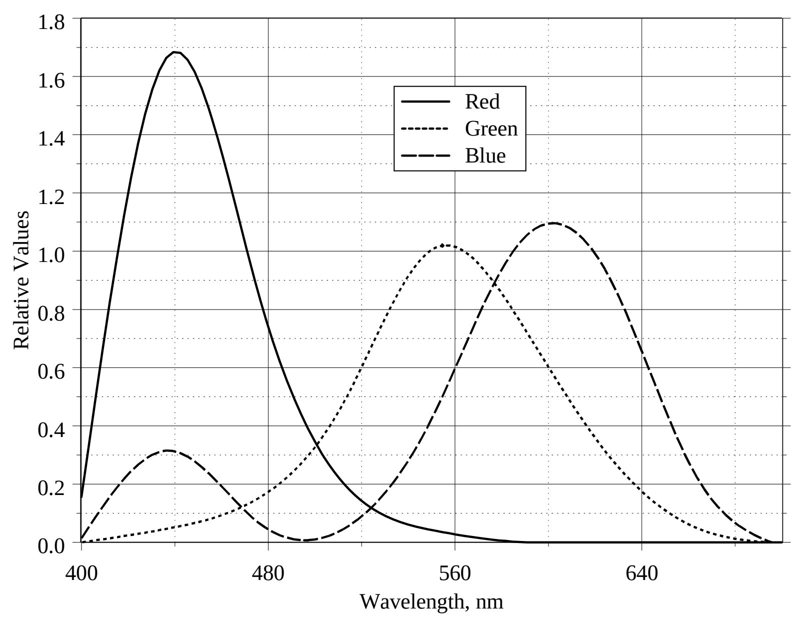


Figure 7-5. Sensitivity of the eyes three cones (simulated).

a light spectrum. After gamma correction (next section), the amplitudes of these three filter outputs represents the **RGB** (Red-Green-Blue) signals that are the basic building blocks of all video signals. As electrical signals, the three colors are usually denoted as E_r, E_g, and E_b—the three output voltage components of the video signal.

Component Video
a composite of the three RGB signals that yields a luminance signal as one of the components.

Another system of signals is the so-called **Component Video** interface which is based on the model of luminance and color differences such that,

$$E_Y = 0.2999E_r + 0.587E_g + 0.114E_b$$
$$E_{BY} = E_b - E_Y$$
$$E_{RY} = E_r - E_Y$$

where E_Y is the luminance of the image, a fundamental quantity. The other two signals are linearly related to the hue and saturation. The two systems, component video and RGB, are completely equivalent in their coverage of the data since there are linearly dependent, but the actual signals are not actually compatible. Either one can be derived from the other but only with some signal processing. We will see that these signals are in the upper mega-hertz region and processing signals at these frequencies is not trivial. Conversion between RGB and Component Video is something that should be avoided if possible.

There are many different color models that one can encounter in the video world, but they are all consistent in that data from one can be uniquely converted into another. To me, RGB is the easiest to understand, but some might disagree. Dealing with three components of a color vector just seems natural and convenient; however, I do not denigrate any of the others because I have a preference for one of them. I will, however, ignore the others and deal strictly with my own preference.

7.3 The Eye

We have learned that color images can be represented as signals in three scalar variables to represent the color and as two dimensional spatial images that obey Fourier Transform relationship. This could also be stated as a three dimensional color vector in a two dimensional spatial map in either the image domain or the spatial frequency domain. I have already talked about the eye from the perspective of its color characteristics, and, now, I will discuss it from the imaging and linearity perspectives.

The point of introducing the spatial frequency definition of an imaging system is so I can discuss the resolution system of the human eye. This turns out to be a critical factor in the design of video systems, since the resolution capabilities basically drive the complexity of the optical systems in the video components. More resolution than the eye can perceive is wasted, but less than the eye can perceive will be judged as poor quality. Matching the resolution of the video system to the human perception of the image is a highly important consideration.

Viewing Ratio
the ratio of viewing distance to image height.

Since we know how to quantify an optical system in terms of its MTF, which we can measure for a know testing aperture, it only remains to find out what the eye's resolution is. This is a widely varying parameter with individuals and age, but that value that is used in the industry is 1.7 (minutes of arc). Since the eye has a constant solid angle resolution (see Figure 9-1 on pg. 158), it will have a MTF that varies with the subject to image distance, denoted the **Viewing Ratio** (VR) which is the ratio of the viewing distance to the picture height. Figure 7-6 on pg. 127 shows the MTF versus TV Lines for three different VRs. As expected for a constant

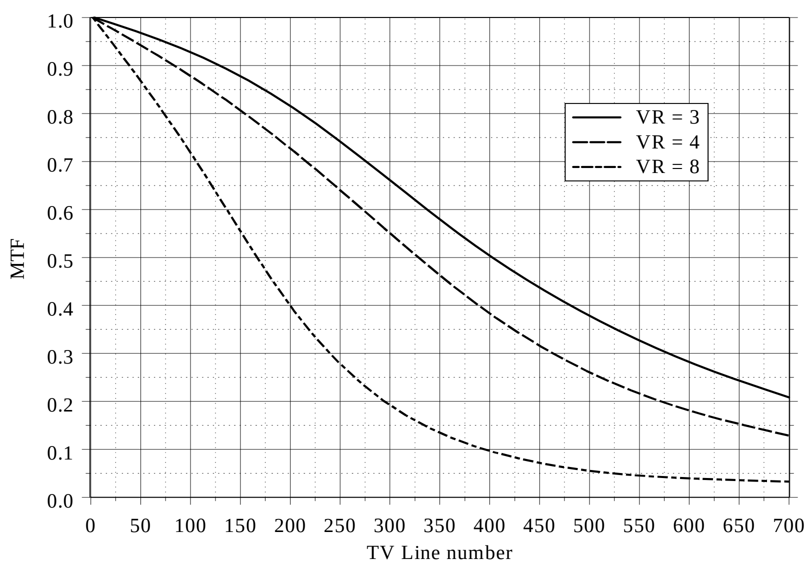


Figure 7-6.
MTF for the eye at viewing ratios of 3, 4 and 8.

angular resolution, the resolution capabilities for the eye vary with distance. I will return to this curve when I talk about video picture resolution.

The last thing to consider when discussing the eye is the logarithmic nature of the eye's perception of light intensity. Just as with sound the eye responds logarithmically to light intensity. This has several direct effects on gray scale perception and underscores the need for a gamma correction to video signals. Since most of these topics are based in the actual equipment and viewing environment design, I will postpone a discussion of them until the next chapter.

I have now shown how the spatial resolution of an image is quantified and how to relate this measure to the capabilities of the optical system, including the eye. I have also shown how color can be defined as a vector in the RGB space and briefly discussed the light perception characteristics of the eye. Now, I can move on to a discussion of video systems.

PART II

VIDEO

CHAPTER 8 Video Distribution Systems

In this chapter, I will discuss the technology that underlies all video systems. I will briefly discuss analog video, but mostly from a legacy standpoint, since analog video is all but dead. Digital Video is the future and the only practical way to do good quality HT. Even large commercial theaters will be digital in the not too distant future.

8.1 Video Image Acquisition and Transmission

At the root of the video design problem is the manner in which video information is acquired, transmitted and displayed. For analog video the acquisition, transmission and storage are each different technologies. For digital video, all the stages are almost virtually identical, being a simple bit stream like the audio bit stream that we talked about in Chapter 3. As I showed in the previous chapter, an image is a two dimensional object that is defined by its spatial color representation. We also saw that the color aspects can be broken down into a three parameter problem. The implications of this simplification is that a color image is nothing more than the composite of three one color images is well known. We can consider a color transmission system to be nothing more than the parallel combination of three distinct images encoded as light intensity.

I also noted in the previous chapter that an image also has a frequency domain representation, just like audio, and that certain aspect of the image problem are better defined in this spatial frequency domain. The Fourier Transform aspects will become important in later sections when we talk about digital video, and where we will see that the Fourier domain is the ideal place to do signal processing of video images. For now, let me just consider the problem of transmitting a video image along a one dimensional medium—like a wire or a radio wave.

Scan lines
a horizontal swatch of an image that is scanned for the light intensity.

It should not be new to anyone that a two dimensional image can be scanned to create a one dimensional data stream which represents the light intensity along progressive lines in the image as shown in Figure 8-1. In this figure it would take 38 lines, called **scan lines** of the width shown to cover the image. If the light intensity is recorded by a sensor that is swept along these lines from left to right (which is arbitrary), then a continuous signal would be created in time (ignoring for the moment the time for the

Figure 8-1. Example photo with a single scan line shown. The picture would require 38 lines of this width to cover its entirety.

sensor to return from the right back to the left hand sides of the image). A "monitor" could then be set up to reproduce the light intensity by the exact same scanning principle and a reproduced image like that shown in Figure 8-2 would be created. There are several important features that can be seen in these figures. First, note that the image has been digitized in the vertical direction by the discrete nature of the scan lines, but it is analog in the horizontal direction, the scan direction. The vertical resolution is easily specified by the number of scan lines—38 **TV Lines** in this example, but the horizontal direction is not as easy to define. In this figure, I have set the horizontal resolution to be the same as the vertical resolution. In the analog video world, the horizontal resolution is set by the maximum rate of change of the data in each scan line, which turns out to be a temporal frequency response issue in the electronics of the video system. More on this later.

TV Lines
a scan line as found in a typical television set.

Viewed closely, the image of Figure 8-2 is recognizable as a face, but whose face it is remains unapparent. If viewed at a sufficiently large Viewing Ratio these two figures should be indistinguishable. From Figure 7-6, I would estimate that the VR required for this to happen would be about 32–64 (an MTF of .1 at 50 TV Lines). Since this image is 3" high, the two figures should be indistinguishable at 32" x 3" = 96" or about 8'–16'. A

Figure 8-2. Reproduced scanned image with 38 lines of vertical resolution and a comparable number (but analog) in the horizontal.

casual experiment indicates a viewing distance of about 24' is required or a VR of about 96. Not too far from my original guess. The discrete horizontal lines are the most evident as a small vertical blurring would reduce the apparent differences in the figures.

By sampling an image in discrete lines, an analog signal can be transmitted in one dimension. There are a few details to consider such as retrace times and overscan (the area around the visible picture), but basically this is how television works in the analog domain.

8.1.a Digital Video

Pixel
the smallest space on a digitally sampled image. They are usually square.

It is now quite easy to move into the completely digital video domain with this same image by simply considering the horizontal scan as being spatially sampled. This result is shown in Figure 8-3. This figure has 50 by 38 discrete squares, or **pixels**. Each pixel is a solid gray of varying intensity. The image could be saved or transmitted with a set of 50x38 = 1,900 numbers. The obvious situation that now arises is how to represent the intensity level, the number, for each pixel.

*Figure 8-3.
A discrete spatial sampled image with resolution of 38 by 50 pixels.*

Gray scale
the range of light intensity from black (0) to white (maximum).

If four bits per pixel were used to represent the **gray scale,** then the image will look like that shown in Figure 8-4. The degradation from gray scale quantization is different than that from resolution limitations, but still a serious consideration in picture quality. Generally 8 bits per pixel is considered to be a minimum and 12 or 16 are more typical. Figure 8-3 is actually 16 bits per pixel.

In order to store the image with 16 bits (2 bytes) per pixel

$$1900 \text{ x } 2 \text{ bytes} = 3800 \text{ bytes} = 4 \text{ kB}.$$

It must be remembered that this is a crude picture that requires about four kB to store. If this picture were a single frame of a motion video with 30 frames per second, then 120 kilo-baud of bandwidth would be required to transmit the video. Clearly, video is even more of a data load than audio for even a reasonable quality level. With typical overheads in transmission, the image in Figure 8-4 would be just about the quality limit for video over the internet.

Figure 8-4.
An image with four bits per pixel to represent the gray scale.

Figure 8-5.
Digital image low pass filtered by typical optical functions.

As a final demonstration, Figure 8-5 shows a low pass filtered digital image. The low pass function comes from the resolution limitation found in all optical playback equipment and it tends to make the image appearance better by reducing the appearance of the sharp edges around the pixels.

8.1.b Flicker

The amount of time from one frame to the next is called the frame rate, and it has a strong visual effect know as flicker—the subjective impression of a pulsating image caused the non-permanence of the image

8.1.c Interlacing

In order to cut down on the speed at which data must be sent, many transmission standards use interlacing. Interlacing is a technique whereby each frame, a single image in a series of images, is broadcast as two fields, each field being composed of alternate lines of a single frame. Thus, lines 1, 3, 5, etc. are transmitted as one field and lines 2, 4, 8, etc. are transmitted as a second frame. For analog television, this cuts the broadcast bandwidth requirements in half for a given image resolution at the cost of only a small amount of increase in flicker. Interlacing is used in many formats of both analog and digital transmission systems. It is important to note that the nature of a computer monitor—images are refreshed from a memory array at various refresh rates—means that computer video systems are never interlaced, even though the data may be transmitted that way.

8.1.d Gamma Correction

It is now time to return to the subject of intensity perception that I glossed over in the last chapter. If I encode the intensity values of each pixel linearly, then I will not have optimized the visual perception of that image. This is because of the ambient light and maximum light capabilities of any video playback system. It is not so true of printed images as I have been showing here.

Consider the log-log relationship between the image brightness and the scene brightness as shown in Figure 8-6. The one to one relationship between log Image and log Scene brightness is shown as the straight line at 45°. The three curves are for log ambient light conditions varying from low

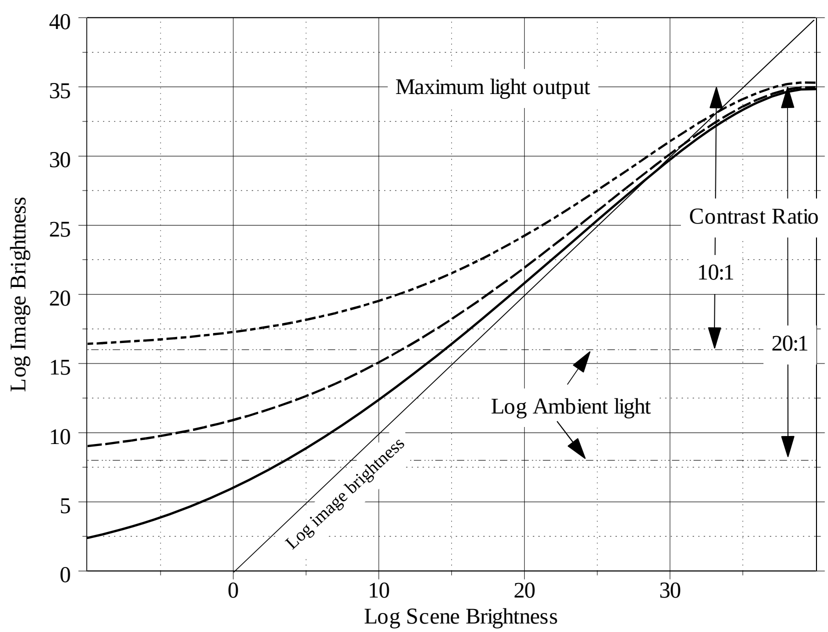


Figure 8-6. The log Image Brightness versus log Scene Brightness for three different ambient light conditions: low ambient, medium ambient and high ambient lighting conditions.

Contrast ratio
the ratio of the lightest to the darkest portions of the image.

Highlight brightness
the light output of the brightest spot in an image.

Gamma
the slope of the image brightness versus scene brightness curve.

Nyquist frequency
the maximum useful frequency for a digitally sampled signal, equal to one-half of the sampling frequency.

to high. The maximum light output of the device sets the upper limit. The **contrast ratio** is defined as the ratio of maximum light output (sometimes called **highlight brightness**) to ambient light. The curves are not straight lines but more S shaped. For best viewing, this effect must be taken into consideration in the playback environment.

Clearly, for maximum contrast ratio the ambient lighting is the limiting factor. It must always be remembered that an ambient lighting condition of zero is not possible since the screen reflects almost all of the light incident on it. It is this later factor that makes the use of rear projection preferred in situations of high ambient lighting. The general slope of the curves in this figure is called the **gamma** value—which was borrowed from the photography field in the days of film.

In order to optimize the gray scale that is present in the image, it must be correctly fit into the available contrast ratio in the environment. This is why it is so important to correctly set the image brightness and contrast for the specific conditions under which the image is viewed. Too high of a contrast will cause the image to appear washed out and fuzzy. Too low and the image will not have as much apparent perceived depth as it would otherwise have.

In some situations, the gamma can be different for different colors and each color must be correctly set for best visual impact. Hopefully, this is not the case (it is difficult to do the gamma for three separate colors) and setting the gray scale gamma should suffice.

8.1.e Aliasing

Based on my discussions in the previous chapter, one might logically ask why the optical systems being used don't simple have flat MTF frequency responses (see Figure 7-3 on pg. 122). The reason is the same as it is for a digital transmission system—aliasing. When any sampled signal contains frequencies more than half of the sampling frequency, called the Nyquist frequency, then spurious artifacts called spatial aliasing are created. This happens spatially as well as temporary in an audio signal. Since a cycle takes two TV Lines to create, the Nyquist frequency is the same as the number of TV Lines.

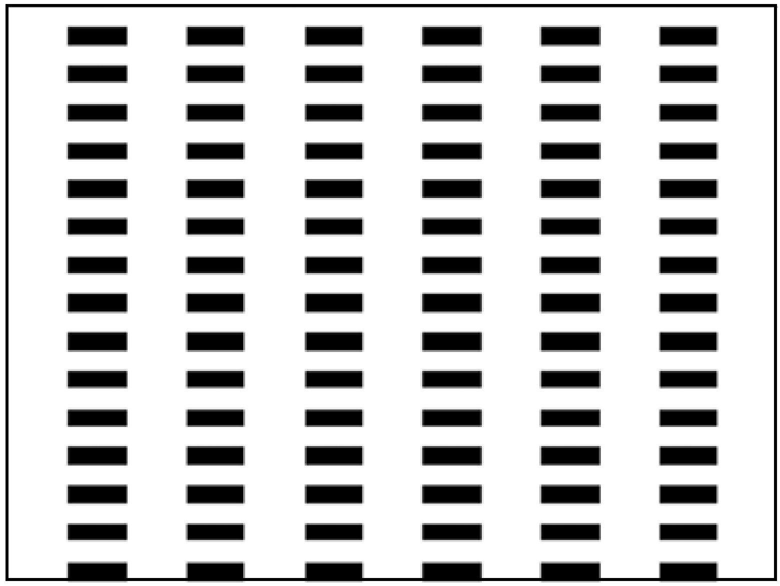


Figure 8-7.
A spatial frequency image with a vertical frequency of 30 TV Lines and a horizontal frequency of about 10 TV Lines.

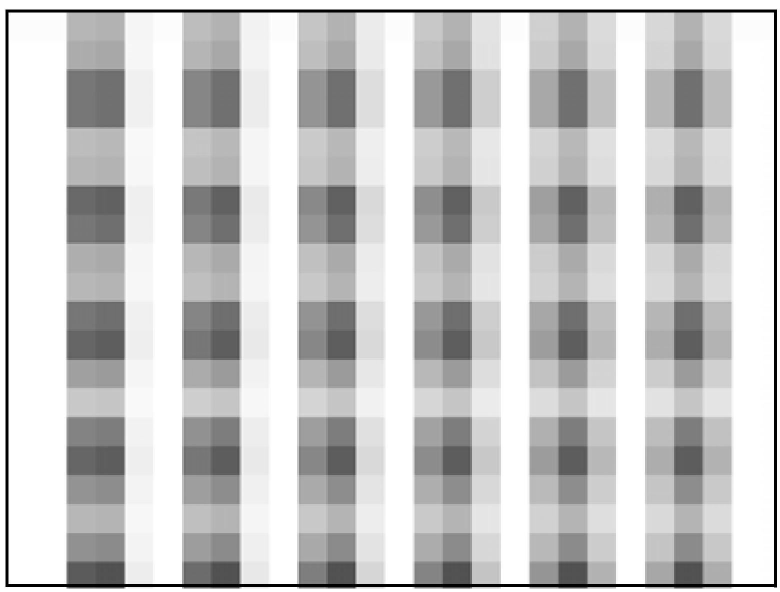


Figure 8-8.
The image of Figure 8-7 sampled at 20 TV Lines. Note that the horizontal line density has been preserved, but the vertical has not. There appears to be only about 5 horizontal lines in the image.

As an example, consider the image shown in Figure 8-7 and its sampled image in Figure 8-8, on the previous page. The spatial sampling frequency of 20 TV Lines is not high enough represent the vertical line rate of 30 TV Lines, but it is for the horizontal rate of 10 TV Lines. The horizontal lines are aliased at the difference in the sample rate and their frequency rate divided by two, i.e. 5 TV Lines, which can clearly be seen in Figure 8-8. The vertical lines are smeared in space by the sampling, but their number is preserved.

Aliasing can be a most disturbing visual effect. I should point out that in video there are two types of aliasing, one due to frame rate and the other due to spatial sample rate. The frame rate aliasing only causes odd effects which are not that annoying, like wagon wheels appearing to run backwards; however, the spatial aliasing can cause annoying modulation effects on fine details in the image, especially if the image moves slowly. Line rate aliasing is often confused with frame rate aliasing, but they are distinctly different things.

Now we can see why it is necessary to have low pass optical systems, even though they are the default, and why it is not advisable to improve the optical capabilities of the capture device (the camera) beyond what its sample rate will allow. Improving the optical image qualities requires a comparable improvement in the entire optical chain. It is impossible to restore an aliased image just as it is impossible to restore an aliased audio signal.

8.2 Television

Television is currently in a transition. Its analog legacy still prevails, but no one would consider using analog television as a serious source material in a premium HT system. I will show why this is and describe the background for HDTV—the next generation of television.

8.2.a Legacy Formats

Legacy television is simply an implementation of the line scanning technique that I defined in the last section. Progressive lines are scanned and the analog signal of intensity levels are transmitted in fields where two fields make up a single image or frame. There are three dominant analog television standards known as NTSC (US and Japan), PAL (Germany, the UK and China) SECAM (France and Russia). Table 8-1 shows a comparison of these formats. The bandwidth is the width of the frequency modulation about the carrier frequency that is required to carry a single channel. In other words, this is the minimum channel spacing. PAL and SECAM have the better resolution but allow fewer channels in a given frequency band allocation.

Table 8-1. Television format comparison

System	Active TV Lines	Frame rate (sec^{-1})	Bandwidth (MHz)
NTSC	480	30 (29.97 color)	4.2
PAL	580	25	5.5
SECAM	580	25	6.0

When talking about color television, there is an aspect of the color signal that is quite different from what we have looked at thus far. This situation developed out of the need for color television to be backward compatible with monochrome television. In a monochrome television, the signal is simply the **luminance** (see pg. 123) versus position. This signal must be transmitted with every color signal as a stand alone quantity. The color information is then carried as a additional single. We know from the previous chapter that the luminance signal can be derived from the three signal E_r, E_g, and E_b. The hue and saturation signals are carried by a higher frequency subcarrier that is not seen by a monochrome television. Since the complete color information is carried on a single channel, it is called **composite video**.

Composite video
the earliest video signal format where all video data was transmitted over a single cable resulting in significant signal interference.

Chrominance
a signal derived from the primary RGB signals that represents a color difference of the image. It is obsolete in three signal systems.

Composite video signals are notoriously bad owing to interference between the luminance and the chrominance signals. **Chrominance** is a signal which is derived from the RGB signals as a vector component. This derivation would be important if S-video were not an obsolete video signal, but since it is I won't elaborate on the details. S-Video is preferred to composite video because it carries the luminance and chrominance signals separately reducing the interference for a much improved picture quality. Good S-Video is the minimum signal quality that is acceptable in a good HT system.

This is about all that I want to say about analog television, since its retirement is only a few years away and its use for HT is quite limited. I will now move on to discuss HDTV and some of the advantages that it has over the current analog implementations.

8.1.a HDTV

The first thing that I must emphasize is that HDTV was created with the ability to utilize a wide variety of formats. HDTV is not a single format, but a whole range of formats and the only common factor is that they are all digital. In fact, some of the lower level formats are simply digital implementations of the old analog formats. They are improved through the use of digital techniques, but the basic resolution is the same as NTSC. Some people, like myself, believe that the term HDTV is being used when simply Digital television (DTV) should be used. Some of the formats are truly high definition while others are not.

Interlaced
the technique of sending only half the picture in a block of data and sending the second half in the next block.

Progressive scan
when each picture is composed of the entire frame data - not interlaced.

A list of the currently used DTV formats is shown in Table 8-2. The last three formats could be called true HDTV while the first three are simply DTV. The frame rates vary extensively from 24 Hz – 60 Hz, but the **interlaced** formats are always about 30 Hz.

Table 8-2. HDTV format comparison

Name	Active lines	Pixels / line	Aspect ratio	Scan method
480i	480	640	4:3	interlaced
480p	480	640	4:3	progressive
480 (16:9)	480	704	16:9	both
720p	720	1280	16:9	progressive
1080i	1080	1920	16:9	interlaced
1080p	1080	1920	16:9	progressive

I want to point out a serious situation that is occurring in the marketplace regarding HDTV. I will also bring this point up later. The term HDTV is highly abused by manufacturers. “HDTV Ready” simply means that you can use an external receiver and connect it to the television—big deal, like “Digital Ready” loudspeakers. “HDTV Compatible” usually means that it has a digital receiver in it and can receive HDTV signals. I have actually seen a television advertised as HDTV that could receive HDTV signals, but the monitor was only capable of 720x540 pixels—it was an LCD display. This is not HDTV resolution—it is DTV. I suggest that you continue to investigate any component that you buy until you find out what its capabilities really are, because, if it is not explicitly stated, then the situation will almost always be less than expected. Otherwise, they would advertise the fact—right?

Now, consider an interesting fact. To receive 1080x1920 pixels at 16bits per pixel times three colors at 60 frames per second is a bit rate of almost 6 gigabits per second. This is an enormous data rate that is not in the foreseeable future. Even the lowest digital television transmission would require 330 megabits per sec or a bandwidth of about 500MHz. This is exclusive of the requirements for audio!

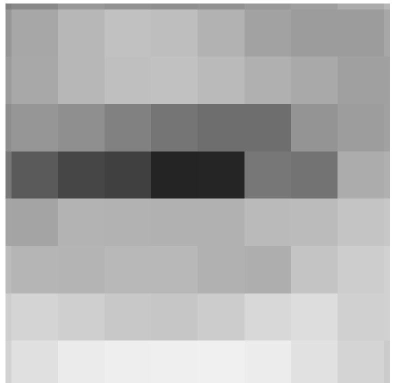
Each television channel is only allocated about 5 MHz, or less than 1% of the bandwidth required to transmit even the lowest quality digital picture without audio. We have seen this situation before and we will find the solution for video is nearly identical to that for audio. In fact, they were developed together, so this is not surprising. What is surprising, perhaps, is that sight and sound have such similar underlying characteristics that we can find techniques for both that turn out to be so similar. Just like audio, the frequency domain is where we will find the technology to do the huge compressions required for reasonable digital video data transmission and storage requirements.

8.2.a Video Compression

Video compression is performed in the spatial frequency domain just as audio compression is performed in the temporal frequency domain. Both systems store and transmit the frequency domain information and then reconstruct the image (spatial or temporal) at the receiving end. The group that came up with this elegant scheme was the Motion Picture Experts Group MPEG. MPEG is now synonymous with digital compression and MP3 stands for MPEG I layer 3 which is the subsection of the MPEG I digital compression technology specification in video that deals with the audio signal.

Consider the digital picture shown in Figure 8-3 on pg. 134. Let take a sub-portion of that picture that is 8 pixels by 8 pixels from some where in

Figure 8-9.
An 8x8 pixel block from the center of Figure 8-3. The DCT is shown as the matrix on the right.



172	-3	4	-2	-1	0	1	0
-22	2	0	-2	0	0	0	0
26	7	-10	0	0	0	-1	1
8	1	-3	-1	1	0	0	0
-6	-3	1	0	-1	-1	1	-1
-12	-3	6	1	0	0	1	0
0	1	0	-1	1	0	-1	0
6	2	-2	0	0	0	-1	0

DCT
the Discrete Cosine Transform, a subset of the Fourier Transform where only cosine terms are used (the Fourier also uses sines).

the center where there is a significant image detail. I have shown this image in Figure 8-9 along with what is called the **Discrete Cosine Transform** (DCT) of this image. I have rounded the numbers to the nearest integer for convenience.

The DCT is the basic building block of all MPEG encoding and decoding because it is an extremely compact simplified version of the Fourier Transform. It is so simple that it can be hard coded into a chip for very fast processing. This later aspect of the DCT is actually a requirement since a thousand or more of these sub images must be processed for each frame. This means that this transform must be done very quickly if it is to be useful. We will come back to this speed issue later.

In the spatial frequency domain, the lowest frequency is the upper left element, which is actually the DC or average value of all of the pixels. The further away from this corner the element lies the higher in spatial frequency that it is. Note that a great many of the higher frequency coefficients are small or zero. In fact, only the upper left triangle has much useful data, the rest are basically insignificant. This means that the number of coefficients to store in the data in the frequency domain is less than half of those in the original image. Also note that the range of the DC term is from 0–255 but the range of the other terms is only half this number. They can, however, be positive or negative which the DC term cannot, so they both cover the same span of numbers.

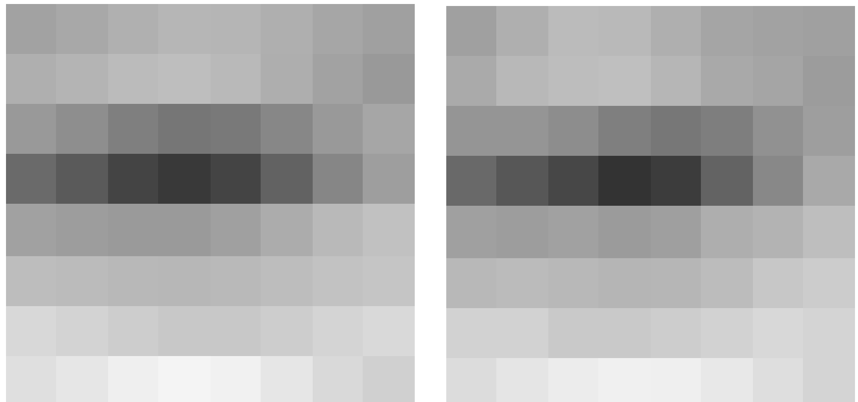


Figure 8-10.
An 8x8 block of pixels where values less than a threshold of 1(right) and 2 (left) have been set to zero. This figure should be compared to Figure 8-9.

Figure 8-10 shows a threshold applied to the spatial frequencies of 1.0 and 2.0. Any coefficient less than these values is set to zero. The effect that this has on the image is shown for 1.0 on the right and 2.0 on the left. The right image has only 26 significant values and the image on the left has only 15 significant values. The image remains similar to the original in both cases, although difference can be seen. At the compression level of 4:1, as in the case shown on the left in Figure 8-10, the image would be nearly indistinguishable from the original one (At typical VRs.)

For regions of the image that are relatively constant in intensity, the entire set of 64 numbers can be reduced to one coefficient.

It also turns out that round off error in quantizing the coefficients in the frequency domain creates a spatial noise that is not highly perceptible and that gets lost in the normal video noise of the image. This is analogous to the masking effect in our hearing system. In practice, a 12–16 bpp (bits per pixel) image can often be reduced to 4 bpp in addition to a reduction in the number of coefficients shown above. Reductions of data of 30:1 or more are often achieved with this technique with little degradation in image quality. It is usual to use more bits for lower frequencies than higher ones.

This is exactly how JPEG compression works for digital cameras on stationary images. In fact, the JPEG techniques have been around far longer than the MPEG ones, since image compression for wire transfer of images was a real necessity in the publishing business. When a clear need for com-

Motion compensation the technique of looking at sequential frames for regions of unchanging or slowly changing DCT blocks.

pression of motion video became apparent, the MPEG group borrowed from JPEG. The major difference between MPEG and JPEG is that MPEG goes on to consider redundancies that occur in the frame to frame implementation of motion—redundancies which can be exploited for further data compression called **motion compensation**.

The MPEG-1 standard was established for the transmission of motion images and sound at a bit rate that was typical for CD's and T1 data lines, about 1.5 Mb/s. MPEG-1 uses MP3 audio compression on mono or stereo channels, and little motion compensation is used. As a product in the marketplace, this technology is called Video CD (VCD), a popular format in Asia, but one that never really caught on in the West. Its implementation is easy and it uses standard CDs as its medium, but its visual and audio quality are poor. VCD is not even as good as broadcast NTSC (under good conditions). therefore, it is not an acceptable format for a HT.

In order to improve on the quality of MPEG-1 the MPEG-2 standard was created. This standard has improvements in these areas

- higher resolution video—more pixels
- higher compression ratios in the algorithm
- higher quality audio with multi channel capability.

Since a higher bit rate up to 15 Mb/s was assumed, MPEG-2 became a vast improvement over MPEG-1 and forms the basis for virtually all video recording and transmission schemes in use today. I will deal with the first two features of MPEG-2 since they are basically the same thing. The capacity for DVD was known at the time of MPEG-2 development and the data transfer rate of about 5 Mb/s was sufficient for a two hour movie. As we will see when I talk about the DVD format, there are a lot of options built into MPEG-2 for adjusting the data rate.

Motion vector the set of two numbers indicating how much to move an object from frame to frame.

The most important considerations for video compression in MPEG-2 has to do with motion compensation. In MPEG-2, only those parts of a frame that have significant changes from frame to frame are actually transmitted (or stored). For instance, if a car were racing across the scene while the background was stationary, then only those blocks of data that contained the moving car would be sent. The background blocks remain the same and are simply reused from frame to frame. Actually, the car itself may not even be transmitted, only its **motion vector**, telling the codec where this car will be in the next frame. Since it is not uncommon to have the background panning slowly, MPEG-2 updates the complete frame every so many frames. This can sometimes be seen in a movie where the background appears to move in discrete stages with an objects motion sweeping smoothly through the scene. This artifact can be very annoying in extreme cases, but can be fixed with a faster refresh rate on the background blocks. The downside is that a faster refresh rate substantially increases the bit rate and the file size for a given movie length. The actual details of the motion compensation algorithm are quite complex, but the important thing to note is that the MPEG-2 algorithm allows for a lot of adjustment of its features. This means that it can be abused resulting in a poor quality reproduction of the original. When done well, however, MPEG-2 can create a very impressive playback.

The audio portion of MPEG-2 has already been discussed in Section 3.2e on pg. 49—Dolby Digital. It is also possible to encode different digital algorithms such as DTS, etc. onto a MPEG-2 file since the video and audio bit streams are separate once they have been de-multiplexer. Dolby's AC-3—Dolby Digital—is by far the most common and is the standard for HDTV.

MPEG-2 forms the basis for all of the HDTV standards but with different bit rates or each format.

As a conclusion to this section, I want to emphasize how important the spatial frequency domain is for the compression techniques that we have seen for both the audio and video signals. From my earliest days, I have been amazed by how simple some complex problems can become in the frequency domain. When I studied optics, I found that the spatial frequency concept was central to optical processing which added to my respect for the Fourier Transform. When I found that the evolving compression techniques were all done in the frequency domain, my respect for the Fourier Transform grew even more. In more advanced work, I have learned that the Fourier Transform is but one of a class of transforms, virtually all of which have applications in audio. This topic is far too complex for inclusion in this book, but does occupy a great deal of the subject matter of my Audio Transducers book for those that are interested.

8.2.b Next Generation Video Compression

Most notably is the MPEG-4 standard. This standard has not yet resulted in anything widely distributed, but it is available in Windows Media Player, version 9.0. This standard and the MPEG-4 working group was most interested in compression schemes which allowed for lower data rates rather than higher resolution. The resolution capability of MPEG-2, as used in HDTV, are sufficient for high quality playback, but the bit-rates of MPEG-2 is too high to be used over the web. MPEG-4 sought to reduce the bandwidth requirements while maintaining the quality. It is too early to tell if MPEG-4 will have the same impact to the marketplace that MPEG-2 has had.

Wavelets
variable width packets of waves whose amplitude rises and then falls.

The next technology on the horizon makes use of a mathematical technique called **wavelets**. Wavelets replace the role of the Discrete Cosine Transform. From what I have seen this technology it is quite appealing. It is capable of bit rate reductions over that of MPEG-2 by about four times. This technology may allow HDTV resolutions with the standard DVD data rate—HD on a DVD! Cool! The first implementation of this technology was in MPEG-4, but only for static pictures. Windows Media Player currently utilizes this capability.

The advantage that wavelets have over the DCT is that DCT has problems with the edges of the 8x8 pixel block because of the nature of the Cosine function. Wavelets go to zero smoothly at the edges and provide for a better block to block transition with sacrificing the compression capability within a block. The actual details of the wavelet techniques is a bit complex and not that important until its usage becomes more widespread. Serious discussion can be found on the web.

Mostly at the research stage, these wavelet techniques are destined to become the next generation in video compression.

8.3 Video Media

At the current time, there are a whole range of different video media formats, but I will concentrate mostly on one, the DVD. I will mention some of the others that are evolving as potential contenders in the HD realm.

8.3.a DVD

By far, the most common form of video media is the DVD. It is actually amazing the speed at which DVD overtook VHS as the format of choice among consumers. It took more than a decade for CD to overtake its predecessors of LPs and Cassette tapes and these later formats are still around today. The DVD has nearly overwhelmed VHS in a matter of a few years, currently widely available on par with VHS. VHS tape still sells at almost the same rate as DVDs, but the player sales have all but died out. Laser disks are already a thing of the past.

DVD was such a demonstrable improvement in quality over VHS that its acceptance was almost instantaneous. For HT, it is an excellent choice, although, as I will show, it is already lacking in resolution capability when compared to the playback system capabilities available today.

The details of the DVD format can get extremely complex. The producer of a DVD has a choice of widescreen implementations and total control of image quality. Sorting this all out can be very confusing. I will discuss what I think are the most important features. Think of the typical DVD resolution as being 720 x 480. This is about the best that can be achieved with the DVD format. This is slightly better than 480i, but DVD is neither full interlaced or progressive scan but a complex mixture of the two which can differ with the DVD setup.

There are several points to be made here. First, DVD is not a true HD format since these formats did not exist when DVD was introduced. There is talk of a HD-DVD standard, but as with most “latest and greatest” (like SACD, etc.) I am always leery of what is actually better and what is just a manufacturers attempt to sell new products. I will not be the first to jump on HD-DVD for fear that competing technologies, like D-VHS may win out in the long run.

The second point is that DVD being a DVD-ROM format is basically a bit stream of computer data. It should come as no surprise that a computer is best equipped to deal with the DVD format in all its gory details than is any fixed hardware platform implementing of the format. That is not to say that good hardware implantations of DVD don’t exist, only that the data processing involved in DVD decoding is ideally suited for a computer and, as such, most efficiently done in that manner.

For example, much was made at one point in time about progressive scan DVD players and the early prices of these units would lead one to believe that this was a real technological breakthrough. The fact is that all computer graphics are progressive scan, because of the memory buffers used for the video, progressive scan DVD payback is trivial on these platforms. On the other hand, a hardware DVD player has to add a large enough video memory buffer to be able to construct this progressive picture. What was in fact difficult to do in a hardware player is automatic in a software player.

Consider also the resolution problem of playing a DVD on a monitor with a different resolution, like the now commonly available resolution of 1024 x 768. Somehow, the native DVD resolution must be up-sampled to this higher resolution. How that is done has a big impact on how good the picture looks. If the image that you want to view has a resolution of 720 x 480, then this image must be expanded by 1.42 horizontally and by 1.6 vertically, neither being a nice round number. Even if this ratio were 2.0, simply placing a median pixel at the average value of its neighbors is far from the best method of **interpolation** (the mathematical term for estimating the value of a function between two given points). This technique is also called **scaling** in video parlance.

Interpolation and scaling
the algorithm for changing the native resolution of an image format to another, usually higher, resolution.

This problem is well know in audio, where one might want to convert the signal from one sample rate to a different rate. Consider the situation shown in Figure 8-11. Here we have a set of samples noted as data points. The solid line shows a linear interpolation of the data. The dotted line shows an interpolated curve using neighboring points to estimate the shape of the curve. If a new sample point were to fall at 5.5, then the difference in the linear and the complex interpolated points would be 3.7 or about a 5% difference. This is equivalent to lowering the bits in the samples to about 3 bits. If a high resolution image is to be maintained through a sample rate conversion, then great care must be taken in the re-sampling process. In video there is also the frame to frame issues with which to contend.

Taps
refers to the number of coefficients in a digital filter.

The re-sampling is normally done as a form of digital filter where the higher the number of **taps** or data points in the interpolation the better the interpolated answer will be. It is almost always difficult to impossible to find out what resolution this interpolation is done at. The point that I do want to make is that this process is exceedingly numerically intensive.

When one considers that to decode a DVD requires an exorbitant amount of calculations in evaluating the DCT in creating the image it takes an extremely fast computer to allow for a complex algorithm to do the data re-sampling properly.

It for these reasons that, at the time of this writing, only the fastest computers running the fastest video cards could do an adequate job of decoding and presenting a DVD in a large playback format such as projection. Even as recent as 2001, the only real option for high quality DVD decoding on a PC was to use a hardware MPEG chip, which did all of the MPEG decoding and presented a 720 x 480 pixel image to the computer. The processor could then do quite a good job of interpolation, since the bulk of the computational overhead was handled in hardware. These early PC DVD players with hardware decoders are still some of the best looking DVD playbacks that I have seen.

I will admit that I have also seen some good software playback of DVDs on a large screen that looked good, but these required more costly high performance computers. The hardware DVD players would run just fine on an inexpensive unit, even an old “hand-me-down”, which was what my first HT computer was. The problem is that these hardware decoder boards are all but gone now due the pressure of the software decoders. To me, an MPEG decoder chip belongs in every PC of the future since MPEG decod-

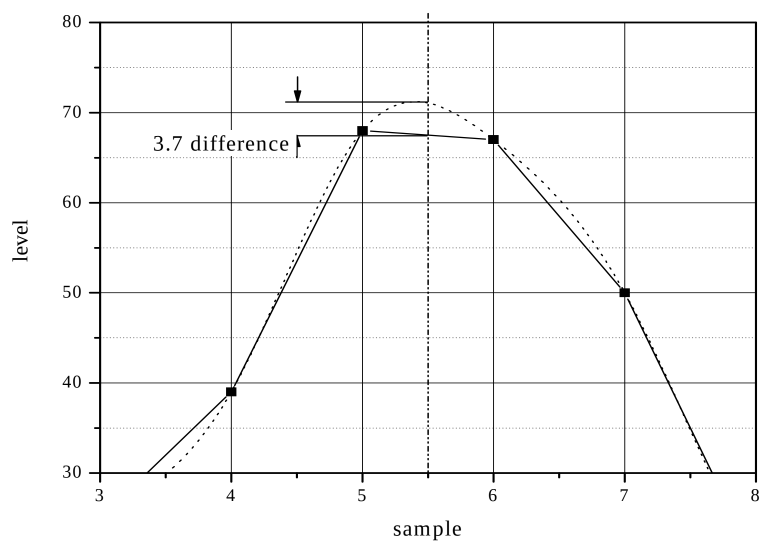


Figure 8-11. Example of an interpolated value and an average value in re-sampled data.

ing is common to virtually all video playback schemes. Hardware decoders can just do a far more efficient job of decoding the basic functions than asking the CPU to perform this task. My current PC actually has three different MPEG decoder chips in it because none of the boards knows enough to share its hardware with the other ones.

DVD represents the defacto standard in video distribution today. Its performance is not only good to excellent, but it is inexpensive to create and easy to distribute. While other formats are appearing on the horizon, DVD will be here for a very long time to come and it will be the center of any HT system. Designing a HT system for DVD then makes a lot of sense. Other formats may give better results, but none of them will even come close to matching the availability of DVD.

8.3.b D-VHS

As I have pointed out, DVD does not have the capability hold HDTV resolutions, that is not for more than about 40–60 minutes per disk. There are numerous technologies for recording HDTV and an HDTV DVD format is being discussed. Since, in its simplest form, HDTV is nothing more than a stream of bits, it is an easy task to simply record this bit stream onto a hard drive. As hard drive capacities skyrocket with little price increase, this is a viable approach. There are PC cards on the market today that do a superb job of recording HDTV signals to a hard drive. The downside is, of course, the size of the files. HDTV takes about 9 GBs for each recorded hour. That fills up even the largest hard drives in a fairly short period of time. A 200 GB would therefore hold about 22 hours of video or lets say 10 two hour movies. About $9 per hour of video at today's prices. This may or may not be a viable way to go for a limited number of recordings, but for a large library it is not workable unless removable drives are used. The other obvious alternative is to have removable media instead of a removable transport—tape comes to mind.

The capacity of tape is actually quite enormous in terms of bits per space. We mostly think of tape as being "old fashioned", but that need not be true. What is old fashioned is analog on tape - not digital on tape. In my opinion, tape may come back as a viable medium for HD video. After all, it is the medium used in the studio for the "master" copy from which all other formats are derived. The current implementation of this idea is D-VHS, a digital VHS tape. It is called “VHS” because it is backward compatible with VHS, i.e. D-VHS machine can play and record standard VHS tapes, but they can also play and record Digital VHS tapes. The D-VHS tapes cannot be played on legacy VHS machines and standard VHS tapes, which are very low cost, cannot be used for HD recordings. If all this sounds confusing that’s because it is.

It is best just to think of D-VHS as a brand new format with new media, but one which has the capability to record HD signals. The tapes look like and are the same size as legacy VHS tapes, but they are not at all the same internally. The D-VHS standard allows for bit rates of 2–28 Mb/s at 6 different rates. The highest is for HD signals, the next (14 Mb/s) for DVD and DTV quality video. The lower rates are for high compression video signals and not of a quality that one would want to view in a quality HT. They can record, however, up to 39 hours of video on a single tape—an impressive feat.

The word on the street is that D-VHS and its theatrical release version D-Theater are comparable, even indistinguishable, from the studio masters. When they are displayed on a high quality HT system, they produce images as good as those seen in a commercial theater. I have seen studio master grade video and it is extremely impressive, and I have had the chance to view D-VHS, but because of the limited software availability the demonstration selections were quite limited. It was hard for me to make a reliable judgement.

One major advantage to D-VHS as being the medium of choice for HD distribution of theatrical performances is that it has a strong copy protection scheme built into the format. This is extremely attractive to the film industry and at the present time D-VHS machines playing D-Theater tapes is the only way that one can get a true HD presentation of a film in their home. From what we will learn in the next section, this may be the only means for viewing the signals in the for seeable future.

Lastly, let me simply say that the future of D-VHS is still uncertain. Quality alone does not insure commercial success and I am not one to risk expensive purchases on uncertain technology. Having gotten burned by my first video format, Beta video tape, I have learned my lesson.

8.3.c Cable and Satellite

As common as these media distribution formats are and as much hype as they are creating around HD broadcasts, their future is not at all certain.

To be sure, cable and satellite digital systems are a vast improvement over analog transmission, but, for the most part, the source material is still NTSC quality. That's probably because the vast majority of customers only have this capability. But the problem goes deeper than that and has to do with bandwidth. As I have noted several times, the bandwidth (bit rate) for HD signals is at least twice that of NTSC, and more like three to four times. For cable, which has a fixed bandwidth dictated by technical issues of signal loss in wire transmission, this is a serious consideration. For each HD cable channel that is added, the cable company must give up two or three normal channels to accommodate it. A major competitive advantage of cable—the number of channels—is thereby quickly reduced by the introduction of HDTV signals.

Satellites have similar issues, per satellite, but the solution here is obvious, use more satellites. My satellite dish, which is HD capable, can receive signals on three satellites. Two are HD capable. But again, the word on the street is that, satellite HD broadcasts of films is done at only DTV quality because of the inability of the service provider to be able to stop copying by the customer. Copying to a computer hard drive would be relatively easy. With its built-in copy protection scheme, this is where the D-Theater D-VHS format has a distinct advantage as a distribution medium.

The future of HD on CATV (Cable TV) and DBS (Direct Broadcast Service) distribution channels is unknown. The DBS companies are making a big push for themselves to be the "home" of HDTV, but they have been doing this for several years now and, to date, the programming in HDTV has been pretty sparse. As I said, there is a real limit to what can be expected for HDTV on CATV systems as long as these systems remain wire based. With optical technology, the cable companies could leapfrog even DBS for capacity and feature content. A single optical cable into a home would provide for all the bandwidth that one is likely to need for the next few decades, but the investment in infrastructure is enormous. Only time will tell what the future will be for HD over these systems.

Finally, let me point out that the terrestrial broadcast companies (your local television stations) are in a pretty good position to implement wide scale programming for HDTV. That's because the HDTV bandwidth is actually only a portion of the channel bandwidth allocated for the old analog systems. That means that as more and more HDTV stations come on line, the capacity for channels actually increases substantially. The original bandwidth allocation was based on antiquated technology capabilities which no longer apply. The lack of a monthly service fee for terrestrial systems means that they have to rely on advertising dollars to a far greater extent than CATV or DBS. Perhaps the ideal distribution for HDTV would be pay television with a monthly service fee structure using terrestrial broadcast as the distribution medium.

I have really only glossed over this extremely important area of video distribution mostly because of its technical complexity. It is, from an engineering standpoint, both fascinating and highly evolved, but it is also extremely complex. The bottom line is that somehow these video signals will get into your home and it is important to understand the basics of the various digital formats since these formats have a large influence on the visual quality of the final performance. The question now is how are we going to handle these signals?

CHAPTER 9 VIDEO DISPLAYS

In the previous chapters, I showed how an AV signal is created and how it gets into your home. Now, I need to discuss how to display these video signals in a manner that is consistent with the nature of the incoming formats.

9.1 Resolution, Screen Size and Viewing Distance

The first thing that must be understood before discussing video display technologies is the intricate relationship between the image resolution, its size and viewing distance. In order to do this, I must draw on the discussion of the resolution of the eye (see "The Eye" on pg. 126) I showed how the eye's spatial frequency response was dependent on the Viewing Ratio (Figure 7-6 on pg. 127). For my present purpose, a better way of looking at this situation is to consider the eye to have a constant **solid angle** of resolution of about 2 seconds of arc (2/60^{ths} of one degree), which is an approximation of the data shown in Figure 7-6. The solid angle approach is convenient for discussing the relationships between screen size, resolution

Solid angle
a circular cone in three dimensional space of a given angle.

and viewing distance. The screen size and the image resolution sets the physical size of a pixel on the screen. The viewing distance sets the eyes ability to resolve an individual pixel.

In Figure 9-1 I have shown the eye, its resolution arc (exaggerated) and the height of a pixel that just matches this resolution. This data can be used to determine the optimum screen resolution, viewing distance or screen size given constraints on either of the other two parameters. Clearly, the ideal is

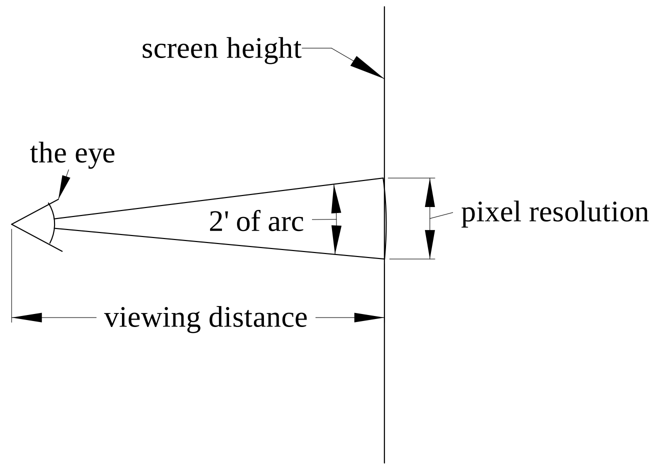


Figure 9-1.
The resolution of the eye showing its relationship to the Viewing Ratio (VR).

to have the eye's resolution exactly match the screen's pixel size (think of eye's resolution as an aperture whose diameter increases in size with distance). The distance at which this occurs for a particular screen configuration is called the **optimum viewing distance**—where the image resolution matches the eye's resolution. Closer distances than this and the eye will perceive the pixilation in the image, further than this and the image resolution is wasted because the eye can't resolve the details.

Optimum viewing distance
the distance at which the eye's resolution aperture matches the screen's pixel size.

Table 9-1 shows the calculated optimum viewing distance for various combinations of screen size and resolutions for both 4:3 and 16:9 formats. This is one of the most useful tables in this book for it can be used to determine screen size, resolution or viewing distance given constraints on any of the others.

For example, in a small room one might have 8' from the seating position to the screen. If the consumer is looking to purchase a new HDTV 60" wide then a resolution of somewhere between 704 pixels and 1280 pixels

Table 9-1. **Optimum Viewing Distance**

4:3												
resolution	640 x 480				800 x 600				1024 x 768			
width (in)	36	52	72	144	36	52	72	144	36	52	72	144
height (in)	27	39	54	108	27	39	54	108	27	39	54	108
pixel size (mm)	1.4	2.1	2.9	5.7	1.1	1.7	2.3	4.6	0.9	1.3	1.8	3.6
optimum viewing distance (ft)	8	12	16	32	6	9	13	26	5	7	10	20
VR (pg. 127)	3.6				2.8				2.2			
16:9												
resolution	704 x 396				1280 x 720				1920 x 1080			
width (in)	32	64	128	160	32	64	128	160	32	64	128	160
height (in)	18	36	72	90	18	36	72	90	18	36	72	90
pixel size (mm)	1.2	2.3	4.6	5.8	0.6	1.3	2.5	3.2	0.4	0.8	1.6	2.1
optimum viewing distance (ft)	7	13	26	32	4	7	14	18	3	5	10	12
VR	4.3				2.3				1.7			

wide is the target. In fact, a 1024x768 screen at this size is ideal. A horizontal resolution of only 800 pixels would not be sufficient.

On the other hand, someone buying a 1280x720 projector when they only have room for a 5' seating distance would only want the screen to be about 40" wide. A larger screen would reveal the actual pixels—a most annoying visual effect.

As a final example, consider a consumer who buys a 704x396 pixel screen that is 40" wide. They would want to sit about 10' back to get the optimum picture from this device. If they could only get about 6' back from the screen, then they wasted their money on the larger screen or the higher resolution since they will not be able to perceive these higher resolutions. Clearly, selecting the proper screen size and image resolution is not simply a matter of looking at it in the showroom. One must consider the specific viewing situation that the unit will be placed into.

One thing that is not evident in Table 9-1 is what I will call the image engulfment effect. This is simply the effect of viewing a very large screen whether the resolution is optimal or not. There is no doubt that bigger screens are more impressive than smaller ones, but this impression can be negated if the image quality is allowed to degrade substantially going to this larger image. So bigger is better if one can maintain an acceptable image quality. Like the audio system, what constitutes an "acceptable image quality" is a subjective matter. How much one is willing to stretch "acceptable" for cost reasons or whatever, becomes a personal decision. At least you have the background to make this decision an informed one.

9.2 Standard Imaging Devices

9.2.a Television

Undoubtedly, standard television is still the most common form for video presentation. They are reliable, yield excellent performance under certain conditions and are relatively inexpensive, but they definitely do have their limitations particularly for HD images.

CRT
cathode ray tube, a glass vacuum tube with an electron gun exciting screen pixel phosphors.

CRT color gun
a cathode ray tube which projects an image beyond the tube for projection onto another surface.

There are basically two types of legacy televisions (ignoring for now the newer LCD panels), direct view using a Cathode Ray Tube (CRT), the traditional television, and rear projection using three color CRT color **guns**. Direct view CRT television's have a real advantage in bright areas because of their inherent brightness and contrast ratio. They can produce excellent color images even in high ambient lighting conditions. The downside is that they really cannot go much larger than about 36" (diagonal) in the direct view format due to the difficulty in producing picture tubes of this size. I will show how this size limitation seriously impacts the higher resolution formats.

From Table 9-1, we can see that a 32" wide television with high resolution would require that one sit ridiculously close to the screen to utilize this resolution—3' for a 1080i signal and 4' for a 720p signal. No one would, at least I wouldn't, sit this close to a television. The current television resolution of 480 scan lines was specifically chosen for a television with a screen of about 20"–30" in size to be viewed at about 6', what the industry thought was the ideal. This assumption is currently invalid, since large televisions (>30") are now the norm. This is what has driven the need for HDTV, a large screen needs this higher resolution to be optimized. A 60" wide set needs a 720–1080 resolution to be optimized for a 6' seating position. To get the most out of these new formats, direct view CRT televisions are simply not big enough to effectively utilize HDTV.

This brings us back to the computer situation. Because people do actually sit very close to a computer monitor, screen resolution has always been a major factor in image quality. It is precisely for this reason that computer display technology has led the industry in resolution and image quality. This is also the reason that computers are, the ideal base component for the entire HD HT—they handle HD resolutions quite effectively.

9.2.b Rear Projection Television

The next technology to come along was rear projection using CRT color guns. Basically, each gun is a individual television tube, but one that projects the image beyond the tube as opposed to on it from the inside. The image is generated by a scanning electron beam identical to a tube television only now there are three of them, red, green, and blue. Each gun produces one of the RGB color signals. These guns project an image on the back side of a translucent screen usually after being reflected off of a mirror. The mirror allows the guns to project vertically creating a substantial reduction in depth requirements, which would otherwise be quite prohibitive. Rear projection systems are not as bright as a direct projection television, owing to the much larger area that the light output must cover. Increasing light output from these guns is not a trivial task. Rear projection systems using CRT guns usually have resolutions that are comparable to that of a standard television although higher resolution systems are now being made.

Three gun systems have one serious problem and that is convergence, which is a problem with any CRT system, but at a much lower level in the one gun systems. The color guns use a scanning method identical to that of a classic television picture tube. This type of system is, if you will recall, is basically a mix of analog and digital. The vertical image is digitized in scan lines, but each line contains analog information. The actual scanning signals used to control the guns location (using deflection coils) are usually analog although some systems use control signals that are created digitally and these are far superior to the analog approaches in terms of stability and ease of convergence.

By its very nature, the scanning system is highly sensitive to component parameter drift which causes the image to shift and/or warp slightly, differently for each gun. Keeping the image of the three guns identical in shape, orientation and location is a major problem and one that I have not found to have been adequately solved (in my limited experience with later generations of this technology).

Usually, (but not always) rear projection televisions have good convergence when they leave the factory. Inevitably however, they will go out of convergence requiring a service call or doing it yourself. (If you have ever seen a typical convergence panel in a television you'd know why you want to avoid doing this yourself.) Redoing convergence is an extremely arduous task and almost invariably the results are not as good as the original. I owned a rear projection television of high quality and found that its convergence was out after a year or two. It became uncorrected after about three or four years. It simply got so bad that neither I nor a professional could get it back to the level of a new set. I swore never to buy a three gun system again.

A friend of mine once commented to me that his rear projection set was so bad on delivery that he had to have it converged immediately. I wonder how long his set will last? I simply cannot recommend this type of imaging system, particularly when one compares this technology to the more modern implementations that I will discuss later.

9.2.c Front Versus Rear Projection

It is not always apparent what the advantages and disadvantages of front and rear projection systems are and how these different implementations affect image quality and contrast. Consider the following.

A front projection system usually needs a screen that is highly reflective to produce enough reflected light to yield adequate highlight brightness (Figure 8-6 on pg. 137). This is where the newer high output projectors come in handy—they don't need a highly reflective screen to get a desirable brightness. A reflective screen will also be a good reflector of the room's ambient light. The combination of images reflected light and ambient reflected light reduces the contrast ratio and in some extreme cases can even make the picture un-viewable. Reducing the ambient light in a front projection system to an absolute minimum is a critical requirement, and an extremely difficult one if sunlight is the offender. But, even in a completely dark room, there is ambient light created by the light reflected off of the screen from the projected image. This situation is completely analogous to sound reverberation. Once the image light has been seen by the viewer (first refection) the remaining light, the reverberant light if you will, goes on to become ambient light. The brighter the picture, the greater the reflected and, hence, ambient light. The contrast ratio then is fixed by the screens and the rooms reflectivities and no amount of projection correction can overcome it. The ambient light in a perfectly dark room from the image reverberation can be quite bright as a visit to a commercial theater will attest.

In a rear projection system one does not want a reflective screen, it needs to be translucent not reflective, so the ambient light situation is vastly improved through rear projection. The downside is that rear projection is only viable in a HT setting when a mirror system is used to reduce the space required behind the screen. These mirror systems have their own set of problems, but that subject is too far afield to get into here. In an ideal HT the imaging system would be rear projection, but done without mirrors. I have seen this done (but not in a HT) and it produces an excellent image.

The obvious advantage of the front projection system is that it takes less room. The space required for the projection is the same space that is occupied by the viewers. The downside is that the projector is also close to the viewers and its appearance and noise can become issues. Moving it far behind the viewers—into the next room for example—is the obvious and recommended procedure.

9.3 Front Projectors

For very large screens, front projection systems offer the only real choice due to constraints presented by every other type of display. There are three major types of projection systems. They are: three color gun cathode ray tube projectors (identical to the rear projection type but only front projection for a larger image), Liquid Crystal Display (LCD) and, most recently, Digital Light Processing (DLP) projectors.

I have already discussed the downfalls of three gun systems and the inherent problems with convergence that plague these projectors in the pervious section on rear projection TVs. For front projection the situation is not much better, perhaps even worse. Since the light source is not a lamp, these systems have trouble producing high levels of light output. That makes them sensitive to ambient light. The longer projection distances for larger images also makes them more sensitive to convergence problems. Ten or twenty years ago these projectors were the only option and some units were made that have absolutely outstanding performance. But by today's standards these projectors where outrageously priced. This was the principle reason that HT were so expensive in the past—the projectors would cost upwards of $30,000. This is simply absurd by today's standards, unless one is interested in the very pinnacle of performance that these highly tweaked units can provide—which I have mentioned is not the focus of this book. When compared to the more modern technologies that have emerged in the last few years, these projectors are pretty much obsolete.

I am sure that I would get an argument from some videophiles about the superiority of CRT projectors in terms of color rendition, contrast ratio, and resolution. It is true that there are some very high quality CRT projectors on the market, but these units are still in a very high price category. For someone who has a high ceiling, a large budget, wants the very best and is willing to live with the inconvenience of complex setup and maintenance, these projectors may be an option. For a budget minded HT these units just do not deliver the value of their competition.

I will mention one option that you may want to consider and that is the purchase of a used CRT projector. These units were built like tanks and are quite reliable. The current prices for competitive projectors has caused the market for these units to collapse and they can be obtained on the used market for reasonable (compared to new ones) prices at web auctions, etc. A friend of mine did this and got a high performance unit that cost just a little more than new DLP projector. He was a video engineer and could do the installation, set-up and maintenance himself—not the usual situation.

9.3.a LCD Projectors

The second type of projector to come onto the market is the LCD projector. This projection system is actually rather old and dates back more than twenty years. It was a long time coming on the market due to reliability issues with the LCD panels that are still its major limitation. In an LCD projector the light source is a single full spectrum bulb which can be of high intensity allowing these projectors to have a substantial increase in light output over the typical three cathode ray tubecolor gun systems. Making the light source separate from the imaging technology is now common practice and one that provides for a substantial improvement in available light.

ANSI
the American National Standards Institute, the government sponsored standards setting organization.

Luminous flux
light energy per unit solid angle measured in Lumens.

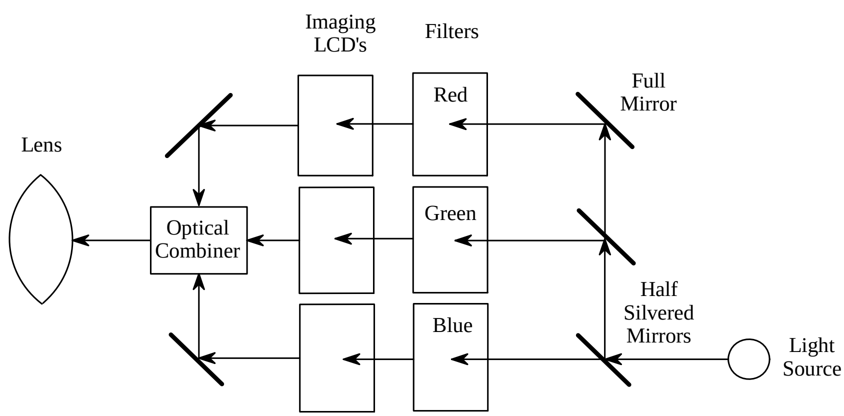
Since I am on the subject of light, I would like to introduce the unit of light output that is used in the market—**ANSI lumens**. The ANSI standard usage is important since other ratings of light output can be fudged, like power amplifier ratings in "peak" watts, etc. This is the unit for the **luminous flux**—the light energy per unit area of illumination, or a fixed solid angle—which means that a lumen will exhibit a decreasing brightness on a surface as the distance to the surface, and hence its area, increases. The lumen is to light what intensity is to sound, and brightness (a subjective term) is to the eye as loudness is to the ear.

The lumen is a linear quantity whose value falls with distance from the source. Being a linear quantity, the eye is sensitive to the log of a lumen ratio, just like dB SPL is the log of an intensity ratio. Since it is an energy quantity—light intensity—the dB calculation uses 10log, not 20log as in the case of dB(SPL)—sound pressure. The lumens, therefore, fall at –3dB per doubling distance not –6dB as a pressure signal does, but the same as the sound intensity. These are a lot of definitions in one paragraph and I hope that you can take the time to sort them all out as you are likely to hear them all at one time or another. You may also see these terms misused such as rating a projectors "brightness" in lumens or referring to a sound pressure level as its intensity. These are subtle but important distinctions.

The dB relationship is important to remember because it means that doubling the lumen output of a display device only results in a single incremental in visual perception. To get a substantial increase in perceived brightness, one must have a substantial multiplication of light output of say ten times. The situation is similar to the loudspeaker problem that I talked about earlier where the maximum sound power output grows substantially with increased dB SPL requirements. Comparing two projectors, one with 1000 lumens and one with 800 lumens, we would find the difference barely perceptible, if at all. A perceived equal increase in brightness would occur for light outputs of 400 lumens to 800 lumens to 1600 lumens.

Returning now to the discussion of LCD projectors, I have shown a schematic of this projector in Figure 9-2. Note that there are two half silvered mirrors. A half silvered mirror is one that allows one-half of the incident light through while the other half is reflected. The final mirror is a full

Figure 9-2. Schematic diagram for three panel LCD projector.



mirror. The astute reader will note that the blue filter will have twice the incident light of the other two and will have to be adjusted with a **neutral density filter** of some other correction. The optical combiner is usually done as a prism system. Finally, the lens focuses the image onto the screen. The lens is often of variable focal length to allow for various screen sizes or projector locations. While this is convenient it is not always necessary if you have a fixed floor plan and deleting the zoom lens usually lowers the cost. Zoom lenses are never as good as fixed lenses, although this is not a big issue in a projector like this, since the lens system is seldom the limiting factor on image quality in these projectors.

Neutral density filter
a glass plate with no color associated with it that absorbs some of the incident light. It will look gray.

Polarization
light is an electro-magnetic wave which is circular in nature, the energy moving between the electric and magnetic fields. Polarized light has one of the axis removed so it is no longer circular but oscillates in a plane.

The LCD panel itself is a common device, but its function can be a bit difficult to understand. It works because of light **polarization**. This can be a difficult concept to understand about a light wave since this sort of thing cannot occur for sound waves. Sound is what is called a scalar wave since the pressure within the wave acts in all directions uniformly. Of course, we know that sound waves do propagate in preferred directions, the waves can have directionality, but this is caused by the momentum of the air particles wanting to continue to move in the direction which they were first excited. The pressure differences within the wave, however, do want to propagate in all directions.

Light waves are different, they are vector waves. The actual wave is composed of both a electric portion and a magnetic portion that act together as a vector. A light wave is actually a rotating vector propagating in space. Quite different from a sound wave in detail, although both waves do act similar in a macroscopic view. The circular nature of a light wave can be broken down into two polarizations, the vertical and the horizontal. This is completely analogous to the vector situation shown in Figure III-1 on pg. 251. For a sound wave, the pressure signal is the real part of this figure, but for a light wave both the real and imaginary parts are required to represent the two polarizations of a light wave.

The liquid crystal has the characteristics that it rotates the polarization of a light wave only if a voltage is applied across it. In an LCD panel there is a grid of electrodes that are individually actuated by a d image signal, a matrix of values, as shown in Figure 9-3. The electrodes are individually actuated in the case of an active matrix display which is the only type used in projectors. Some older computer displays used a different technique for

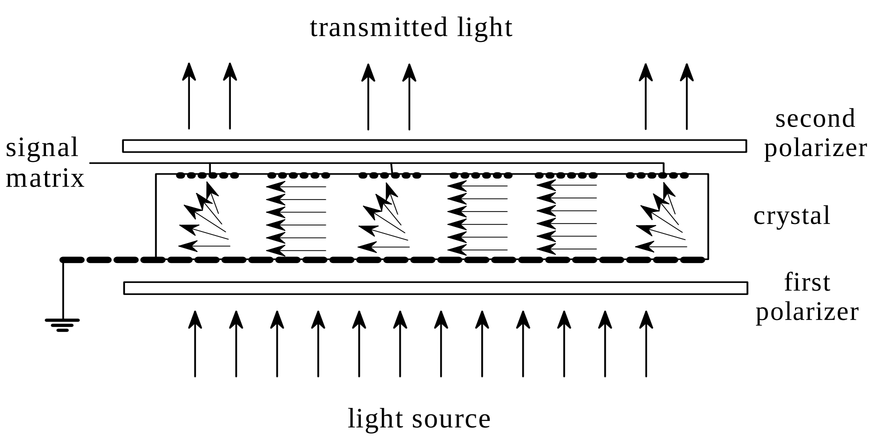


Figure 9-3. Schematic representation of an LCD panel. The liquid crystal is in the center with electrodes on each side, one grounded. Light is transmitted only where an electrode is positively charged.

Polarizing filter
a glass plate optical filter that only passes on polarization of light. Sunglasses often use this type of filter.

selecting pixels—scanning by rows and columns, but this mode is obsolete due to its poor performance.

Non-polarized light is sent though a first **polarizing filter** leaving as polarized light wave—one of the axes is stripped off. If an electrode is electrified then the light wave is rotated by the presence of the electric field across the liquid crystal as it traverses through the liquid. The amount of electric field applied determines the amount of rotation of the polarized light. Hence, the light can be rotated from 0° to 90° by applying different levels of voltage to the electrode. A second polarizer which is orthogonal (at right angles) to the first is placed on the opposite side of the panel.

If the light is not rotated by the crystal, then is will not be able to traverse the second filter due to the fact that the second filter is at right angles to the first. It is also possible to have the two polarizers parallel, in which case a positive voltage turns off the pixel rather than on. If the polarized light has been rotated then it will be transmitted by an amount that varies with the angle of rotation and hence the voltage applied to the electrode. In this way, pixels can be turned on and off with a continuous range of light transmission for each electrode location, or pixel. The image is formed by the transmission of these pixels of light.

The separated color images are then recombined in an optical combiner which is usually made of prisms. The three color image is then focused on the screen by an objective lens placed in the optical path.

Consider what happens to the light that is not transmitted through an LCD, it is absorbed by the polarizing filters, liquid crystal, etc. For a high powered optical system that is a lot of light and a lot of heat generation within the LCD itself. This is exactly the problem that has plagued these projectors for so many years, the LCD's would get very hot and failures would be exacerbated by the severe thermal cycling that results. Even if the LCD does not fail, the word on the street is that LCD panels loose contrast at a rapid rate making them less than desirable in a matter of only a few years. As prices of these projectors fall, replacing them every other year or so may actually be an option for some of the low end units, but it is an expensive option for the higher priced ones.

DLP
Digital Light Processing, a technology where the light is control by micro-mirrors actuated by digital signals.

MEMS
Micro-Machining where microscopic machines are fabricated on silicon wafers in the same manner as IC chips.

9.3.b Digital Light Processing

One of the most appealing technologies on the market today has to be **Digital Light Processing (DLP)**. DLP is a trademark of Texas Instruments and is a **Micro-Machining (MEMS)** technology. MEMS is where a miniature machine is fabricated on a silicon chip in a manner similar to that use for Integrated Circuits. In the case of DLP, the machine is a cantilevered mirror whose orientation can be changed rapidly by a control signal applied to the device. An individual mirror is shown schematically in Figure 9-4. Each pixel contains one of these miniature devices. A bias voltage is applied to the mirror and positive and negative voltages are applied to the electrodes resulting in a push-pull arrangement of electrostatic force on the cantilever yoke. One polarity of voltages tilts the mirror in one direction and the other polarity tilts it the other way—one state is “on” and one state is “off”.

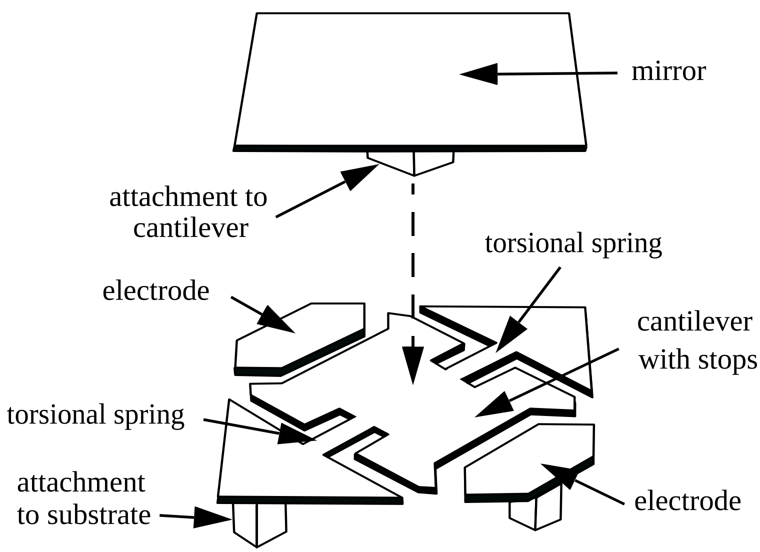


Figure 9-4. Schematic of MEMS cantilever mirror system on DLP chip.

Figure 9-5 on pg. 170 shows how these on and off states can be used to transmit pixelated information regarding the state of the mirror array. There is an array of these mirror devices built into a single chip of about .5"–.9" diagonal. The mirror can switch at rates up to thousands of times per second. By modulating the light being reflected an almost continuous gray scale can be created.

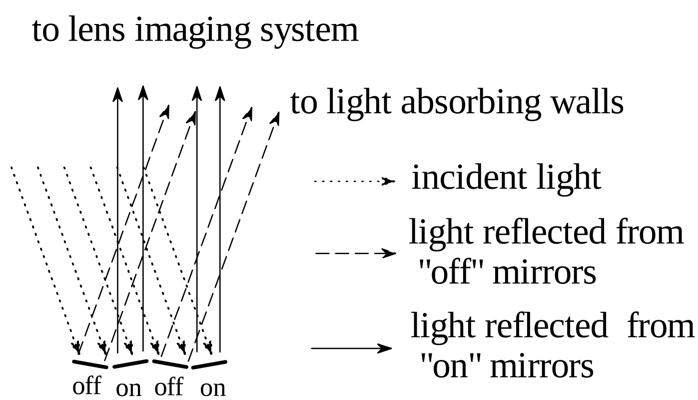


Figure 9-5. Schematic drawing of the light paths for "on" and "off" DLP mirrors.

To get color, the light source is passed through a rotating color wheel with segments that are red, green and blue and sometimes black. This wheel is synchronized to the image display such that the red image is formed during the period when the light beam is passing through the red position of the wheel, the green image the green portion, etc. In some projectors this color wheel has black or white segments interspersed between the color portions. This is to give either brighter whites or deeper blacks depending on the application. For HT, it is most desirable to have deeper blacks since brightness is seldom a problem, so most HT projects have black sections between the color ones. This is just about the only difference in the HT projectors and the computer presentation projectors—given the same aspect ratio.

DLP projectors have several inherent advantages. The first is that they reflect almost all of the light incident upon them so that they have to dissipate very little heat. This is a major reliability advantage for these devices. The second is that they have a better packing density than LCD devices which creates a denser, more vivid image than LCD.

DLP's seem to be sweeping the market at the current time and could become the commercial theater projector of the future allowing direct digital distribution of films resulting in a dramatic reduction in the cost of movie distribution. Whatever type of HT imaging device you select, make sure that you consider DLP as it is an important advance in imaging technology.

LCOS
Liquid Crystal on Silicon is a technology which has the LCD fabricated on a reflective silicon backing which allows for better heat dissipation, cooler operation and better reliability.

9.3.c LCOS

Liquid Crystal On Silicon (LCOS) is a relatively new addition to the array of imaging technologies available. It was developed to alleviate some of the pitfalls of LCD, mostly the heat problem. Basically, LCOS is a reflective system, like DLP, where the image is created by a pixel being either reflected or not reflected off of a reflective surface. Like normal LCD displays, there are two polarizing filters and a layer of Twisted Nematic Liquid Crystal on top of a silicon substrate. Having a reflective surface, as opposed to two surfaces which have to be kept optically clear, this device can be better cooled from the back than a normal LCD which has a limited ability to cool itself since it must be kept optically clear and can only be cooled from the edges. LCOS is relatively new and product availability is limited. Long term commercial viability is an unknown.

9.3.d Comparisons

Table 9-2 shows a comparison of the three technologies that are expected to compete for the marketplace in the coming years. I have shown the maximum resolution for products currently on the market and it should be noted that LCD has the advantage here because it has been around the longest. It is expected that DLP and LCoS will both have these resolution in the next few years.

Table 9-2. Comparison of Front Projection Systems

Performance Parameter	LCD	DLP	LCoS
Maximum Resolution (currently)	1366x768	1280x720	1400x1050
ANSI Lumens	2200	2400	1500
Contrast Ratio	900:1	3000:1	800:1
Fill Ratio	65%	85%	93%

Fill Ratio
the percentage of each pixel that contributes to the image. Larger percentages are the most desirable. A computer monitor would have a value near 100%.

ANSI Lumens was defined on pg. 165 and contrast ratio was defined on pg. 138. The **fill ratio** is the percentage of a pixel that is actually capable of projecting light. It is an important factor for the clarity of the image.

9.2.a The Room

As in audio reproduction, in video reproduction the room matters. The two most important factors are ambient light (which I have already talked about) and reflected light control. For a front projected image to look best, ambient light must be eliminated, this includes outside light from windows and general illumination in the room. For casual television viewing or for watching bright programs like sporting events, where image quality is not paramount, some ambient light is not that detrimental. With today's high contrast digital projectors; light reflected from the screen, walls, ceiling, floors and furnishings and then back to the screen will make the image look washed out and not as natural as it could. Darker, flatter, colors for the room decoration will minimize the reflected light.

Throw distance
the distance required between the projector and the screen. Zoom lenses make this distance variable.

The size and shape of the room is also important. Projectors require a minimum distance between the projector and the screen. This distance is called the **throw distance**, it is often expressed as a ratio of screen width to distance from the screen. For example, to project a 100" wide image with a projector that has a throw distance of 1.5, the projector must be mounted 150" (1.5x100") back from the screen. Zoom lenses give some flexibility in projector location. Consider the required throw distance and the size of your room when choosing a projector.

Letterbox
the image format for an image that is wider than the screen where there are black bars on the top and bottom of the image.

Light engine
the specific mode of light control for creating the image, i.e. LCD, DLP, etc.

Masking
the technique of absorbing the light in a letterbox situation to improve the perceived contrast.

When projecting a widescreen image at 1.85:1 or 2.35:1 aspect ratios onto a a 4:3 aspect ratio screen, or a 2.35:1 aspect ratio image onto a 16:9 screen, black **letterbox** bars will be present above and below the image. These bars will not be fully "black", but instead will appear dark grey due to reflected light from the room, light scattered in the optical path, and light that "leaks" through the projectors **light engine**. Covering or **masking** these gray bars with a highly absorbent material like black velvet, velveteen or Velux will improve the perceived contrast of the image. These masking bars can be made as removable panels and used when necessary.

Bias lighting
adding some background light to reduce eye fatigue.

For critical viewing, all displays look their best in a completely dark room. Watching a small bright display in a darkened room can cause eye strain as the iris is repeatedly called upon to open and close between bright and dark scenes. This effect is most pronounced with smaller screen sizes and viewing angles (large VRs). An effective remedy is to illuminate the area behind the screen, a method called **bias lighting**. Bias lighting prevents the iris from opening and closing during scene changes and improves visual contrast. Correct bias lighting is accomplished when the illumination level around the display corresponds to 10% of the displays peak white value. Bias lighting also works best when the area around the display is a neutral gray. It can also be employed behind screens used with relatively high contrast ratio projection systems to improve the apparent contrast of the image.

9.2.b Video Signal Processing and Transmission

Scaling and signal processing can be a mysterious and confusing part of the complete projection system, and it is often misunderstood. The basic job of a scaler/processor is to take the video signal from the source and change it so that the projector gets a signal that is a perfect match to its native capabilities (pg. 150).

Digital projectors, like LCD and DLP, present two challenges to video reproduction. First, they have fixed display resolution determined by the resolution of the imaging device and second, the display is inherently progressive scan instead of interlaced like direct view single tube televisions. The progressive display issue is solved by de-interlacing the video signal by taking two interlaced fields and making a full frame out of them. The full frame is then enlarged, or scaled, to fit the projectors display resolution. Digital projectors that accept composite, S-video, and Component Video signals do this de-interlacing and scaling internally, but many digital projectors were designed for the business market and do not do a good job of this image transformation. There are exceptions; some HT specific projectors are designed with top quality processors built into them. We will see that a PC can also do a good job at this task.

Three gun CRT projectors and rear projection CRT televisions do not have fixed display resolutions and are capable of displaying an interlaced signal. A standard resolution (480i) signal will show visible scan lines on a CRT projector. The challenge with CRT projectors is again to de-interlace and scale the image so that these lines are minimized and a smooth image is displayed. This is sometimes called

Processing can be as simple as hooking up your DVD player to your digital processor and letting the internal processor do the job, or as complex as an expensive external processor. If the projector has a good quality processor, you are set, but if you are using a business type digital processor or a CRT projector you will need an outboard processor.

Line doubler
image processing for CRT displays that de-interlaces and improves image resolution by adding interpolated lines between the existing lines.

Commercial processors vary form relatively inexpensive **line doublers**. which perform the de-interlacing function without any scaling, to increasingly complex processors with multiple inputs, and variable scaling rates and output rates.

As I will show in Chapter 12, the HTPC is a very attractive option since they can do an excellent job of the video processing functions. It allows for maximum flexibility in setting output resolutions.

HDTV generally does not require any processing, although there are some projectors (9" CRT's) that can benefit by going from 1080i to 1080p.

9.3 The Screen

With any projection system the screen is an important aspect of the final installation. One can spend a lot of money on a screen, but I implore you to consider this expense from a practical point of view. There are really only a few aspects of a screen that are important—the rest, like motorization, or drop down, are conveniences.

I virtually always recommend a fixed screen. This makes keeping it flat relatively easy and eliminates a lot of necessary expense on motors, remote control, etc. It's not as “cool”, but it is a whole lot cheaper. In a dedicated HT, the screen always being visible is not an issue any more than it is in a commercial theater. The screen then is, basically, the fabric and its mounting—that's about it.

I cannot overstress the importance of flatness. Any deviation from flat will be quite apparent on a moving image. This is the main reason that a wall is not a good screen, because they are hardly ever flat. If you intend to use a wall, cover it with panelling like white Abitibi, sprayed with a matte spray to reduce its reflectivity (or serious hot spots will occur) or a flat gray for increased contrast. I have used this technique and it works alright, but it does have a flaw in that the speakers cannot be placed behind the screen.

This brings us to an important point about screens and that is their sound transmission characteristics. I strongly recommend placing the speakers behind the screen as this is really the only way to get the kind of audio—visual image matching that is intended in the movie sound track. For a large screen, where is one going to put a large center channel speaker if not behind the screen? If the center channel is not behind the screen, then it will not have the center sound image where it should be—directly centered on the visual image. That's the way it is mixed and that's the way it should be played back. I will come back to this point in the construction section.

As far as sound goes, a little perforation goes a long way. Basically, for most perforated screens there will be only a few dB loss of the high end that's all. I don't know of any studies of the effects of a screen on the polar response of a loudspeaker placed behind it—presumably this is because there is really only a negligible effect. Perforated screens are available form nearly all of the current screen manufacturers in a variety of screen finishes.

Screen gain
the ratio of the light that is reflected off of a screen to the light that is incident on it.

One of the more important parameters of a screen is its gain. **Screen gain** refers to the amount of light reflected from the screen as a ratio of incident light. In years past, this value was a big consideration. Today, it is much simpler, for there are basically only three choices .7, 1.0 or 1.3 for HT. A gain of more than one means that more light is reflected in a particular direction than a normally incident light would cause to be reflected. This can a happen with some finishes where beads are used to collect and enhance the light's refection back to the audience. In the past, when projectors were starved for light output, the audience needed all of the screen gain that they could get. The problem with this gain is that it creates a hot spot on the image which is a most annoying effect when viewing a movie.

A gain of one is what we usually see—a plain flat white screen. More recently on the market is a screen that is light gray in color with a gain of .7. This is the screen that I recommend because it can help to bring back some of the deep blacks that are hard to create with front projection (Figure 9.2.c). The contrast and color rendition of these screens is really quite good and they go a long way toward making front projection the mode of choice. Recently some gray screens have been offered with a gain of 1.0, but .7 is the more common.

There are also screens on the market which are gray and have a gain of 1.3, but I have no experience with these screens other than to know that they do exist. This would be a good option if more screen brightness is required (as I will describe next) and you want the higher contrast that a gray screen can provide.

When choosing a screen size and screen gain, the system should be designed to ensure sufficient brightness. “Sufficiently bright” has been defined by the Society of Motion Picture and Television Engineers (SMPTE) in standard 196M as 12–22 foot-lamberts (41–75 cd/m^2), a nominal goal of 16 foot-lamberts is often used. Screen brightness, in foot-lamberts, is simply the product of the amount of light energy produced by the projector, in lumens, and the screen gain divided by the area of the screen measured in square feet.

Beyond choosing the gain, which, as I said, for modern projectors, will virtually always be around .7–1.0, there remains only the choice of perforated or not. The first choice is really a choice of color, gray or white and the second choice depends on whether or not the speakers will be placed behind the screen. That is just about all there is to choosing a screen. Mounting it is the only other aspect of concern and this varies so widely with application that I will delay this discussion until the application section.

I want to mention one experience that I had when building my own HT since it was enlightening for me. After I had received my projector, I was still several weeks away from delivery of my screen fabric. Being impatient, I tried using a wall, but, as I said, the flatness was a real problem. I then used a pure white 4'x8' enameled panel (Abitibi) as the screen, which corrected the flatness problem, but it had a real hotspot problem. A can of artists matte lacquer solved the hotspot problem, but I did not have a good place for the speakers since the design had them behind the screen. I finally went to a “white goods” sale and bought an inexpensive (but decent quality) bed sheet on sale. I framed this material like one would frame the fabric for a painting, and then I sprayed it with a fine mist of water. When it dried it was tight as a drum—perfectly flat—and was acoustically transparent.

We used this screen for several weeks and I had few complaints about it. In some rare cases the texture of the fabric was apparent on bright scenes, but, for the most part, this solution worked fine. The screen material, when it finally arrived and was installed, was only marginally better, yet it was a slight improvement. Whether it was worth more than ten times the cost is debatable and in my wife’s opinion—not! My new screen was not even an expensive one, but, actually, the lowest cost screen fabric that I could find.

The bed sheet was not gray, although one might be able to dye it. I have no experience with dying fabric, but I do know that it would have to be extremely uniform in color to be acceptable. I am not really recommending this approach in any but the most tightly constrained budgets, but it is something to think about. I am saying that, based on this experience, it is really hard to justify anything but the lowest cost screen material.

9.4 Other Displays

Probably the most interesting new displays are the plasma screens that are proliferating the marketplace. These screens yield an extremely high quality image—but at a price. The technology is one where an electric signal causes a plasma to emit ultraviolet light which excites a phosphor, like on a CRT picture screen, thereby emitting light of a certain color. These units can have high resolutions and brightness and work well in poor light situations (too much ambient light). But as a large screen format (>40"–50"), they are simply too expensive to be taken seriously. This is not where I would spend my money. Also, they cannot be made sound transparent. Like it or not, movies will always be made for commercial theaters where the speakers are behind the screen. To me, this is the only location that is acceptable for the center channel and so an acoustically transparent screen is dictated.

9.5 Conclusion

In concluding this chapter, I want to say that while there is a wide variety of possible solutions to the imaging problem, my recommendations are pretty specific. Direct viewing does not allow for a screen large enough to give that “big screen” theater effect, no matter what technology one is talking about. A projector, therefore, is really called for. In the projector market, at the present, DLP is the winning technology for the moment and may be the winner in the long term. Front projection is the only real option due to installation issues and the fact that only front projection allows for the speakers to be behind the screen. A gray screen makes the front projection image quality quite comparable to the rear projection one and should always be used in a HT.

All in all, my recommendation is simple, a front projection DLP system using a gray perforated screen. The size and resolution depend upon the specifics of the installation.

PART III

DESIGN AND INSTALLATION

CHAPTER 10 HOME THEATER DESIGN PRINCIPLES

It is now time to assemble all of the various disconnected pieces of information that are strewn about this book into a cohesive design task. That task will be to design and construct a HT—one where the goal is to recreate the commercial theater experience in the home.

10.1 Basic Principles

I want to discuss some basic principles that I will be using in my discussions about the HT design. I have touched on many of these before. To me the theater experience is one of feeling immersed in the film. This experience is far more enveloping than simply watching a movie on television. I think that everyone knows what I am talking about here. Watching *Star Wars* on a 21" television just cannot create the same effect as seeing it in the theater. I think that some will doubt that it is possible to recreate the theater experience at home and, to be sure, there are some aspects that the home just cannot create. In the commercial theater the huge screen towers over you, and there is a feeling of being in an event with a multitude of other people. These are things that a HT cannot recreate and they can be important aspects, true, but not all important. In my experience, they are not that important at all. However, I have come across people who, when presented with a truly high performance HT are still under-whelmed. They are clearly looking for something else in the movie experience than just the presentation.

One other aspect of HT that cannot match the commercial theater is that (currently) movies cannot be seen when they first come out. One must wait until they are available through some transportable medium like DVD or Pay-Per-View. This situation may change in the future with high speed internet distribution or pay-per-view releases concurrent with the theater, but such is not the case at the present time. There is no doubt that seeing a movie on its initial release has its attractions.

Then there are those aspects of the commercial theater experience that are happily missing, like the high ticket prices (getting higher all the time), waiting in long lines, fixed starting times and uninterruptable performances. I won't even mention the annoyance of people who won't stop talking, or spill their pop or popcorn on you, or… I want to mention the most annoying experience that I ever had in a commercial theater—one that would not occur in a HT.

I am a fan of all types of film and I enjoy older classic films even though they cannot compare technically to the newer ones. They can still be great films. I had gone to a local theater to see *Roshomon*, a film by the renown Japanese film director Akira Kurosowa. Having seen both Japanese and Chinese opera in their native lands, I understand the Asian approach to acting but most westerners don't. Asian acting is often overly dramatized in order to make a specific point or convey a deep emotion. *Roshomon* was typical in this regard. It was an intense character study of a very unpleasant event and tended to make one very uncomfortable viewing it—intentionally so. While watching the film, some of the audience found the acting humorous and started to laugh at it. The laughter, once started, began to build and build until one would have thought that they were seeing a comedy—which it most certainly was not. For me, the entire experience was ruined and I have only been to a public theater once since that experience. After we did return to a theater, my wife and I both commented that we did not miss it. Later, I saw *Roshomon* in my own HT and enjoyed it immensely.

I have only been to a commercial theater the one time in more than two years and, as I said, I don't miss it at all. With the money I save on tickets and snacks, I estimate that I've paid for my own theater in that time.

In my experience, a HT can equal a commercial theater in impact. Yet, HTs are most likely to be appreciated by a movie buff. If television is your primary interest, then an elaborate setup might not be worthwhile. I have not seen many television programs that really have the AV impact that movies have, but that is not surprising. Television is, for the most part, viewed on smallish sets with tiny speakers so there is little point in creating sound effects that will only be highly distorted when played through these small sets.

I'm not a big sports fan, but the events that I have watched on my HD HT have been impressive—far more so than a small television could ever provide. One visitor to my home was only reasonably impressed with my setup until he saw a football game in HDTV on my big screen. That blew him away! He said that he'd never seen anything like that before and had no idea that such quality was even possible. Clearly, different people are looking for different things in a HT. To me, achieving a first rate capability for film will allow for the same experience for all other program material—so my main focus is on film reproduction.

10.1.a Location, Location, Location

The first step in any theater design is choosing the physical location of the room. I always find it disappointing when people simply make the assumption that the HT needs to be in the family room, or living room, etc. I think it a very bad idea not to at least consider alternate locations to see if one is actually better than another. To me watching program material on a HT is not a casual event. One does not plop down in a HT like I build and put on the television to view the news. It actually takes significant time for the projector to come on, to get the program source running, etc. I suggest a mid-quality television be placed in the family room for casual watching and putting the HT elsewhere. That's what I have done and it works great for me. I usually don't bother to watch normal television on the HT, it just doesn't make that big of an impact. The quality is not good enough to fulfill my expectations. That is excluding HDTV that is. If only there were actually HDTV programs available. Maybe someday.

Then, there are the kids. They just like the television on. They don't watch it for they are always doing something else, but still they want the television on. That would be a very expensive proposition for a projection system where bulbs have a limited life span and cost as much as a television to replace. One thing that one most definitely does not want to do with a projection system is leave it on when it is not in used. I see using the HT as a deliberate event that warrants moving to a specific location to enjoy, which may be well away from the family room. To me multi-purpose rooms end up not being very good at any of their intended purposes. Every "purpose" is compromised.

There are several things to consider in selecting a location. The ambient conditions that will have to be dealt with, the ambient noise levels and the proximity to possible offended parties, things like bedrooms. Obviously, selecting a location that already has low levels of the two ambient level items is a good start, since reducing them is a major task in any design. It is usually easier to find a location that has low light levels than to try and create one in a high ambient light location. In fact, these dark locations aren't generally desirable for any other purpose. Both of my theaters have been in the corners of a basement. One had a window, and the other didn't. The window was never an asset and I would have preferred that it not be there.

If there must be a window in the HT, then consider the window as unusable and seal it up as tightly as possible. The ability to open and close the window shades will lead to either a very expensive solution, or a major light pollution problem. As I have discussed, ambient light destroys the contrast ratio of a projection system very quickly and a poorly darkened room will be found to be useless for viewing during the day. Sooner or later, this will be a problem. Room darkening shades "reduce" the light level, but they simply don't make it dark—pitch black like a room with no windows. Most of your projection quality will be thrown away by even the smallest light leakage around a window. And if there are several windows, or large ones, you can forget it. You are just not going to get the room dark enough so that it will not degrade the image. Most women find the idea of an expensive room with no windows repellent. Then, painting it dark gray doesn't improve the situation any. You simply have to keep in mind what the purpose of this room is—it is not an interior decorating showplace.

If you simply cannot find a location without light pollution, then resolve yourself to the fact that daytime viewing is probably not going to be optimal, or maybe even possible. Depending on location and time of year, this may or may not be a problem. At my location in June the sun does not set until 9:30 or 10:00 p.m. For all practical purposes these are the entire waking hours. The sun's intensity does decline in the evening hours, but it can still be a problem. In the winter, when the sun sets at 5 p.m., ambient light is mostly not a issue. One thing that I will strongly suggest is that you take the ambient light situation seriously. It can be a major problem to correct after the fact and, unfortunately, the solutions usually have many drawbacks.

The next consideration is the ambient noise level. This is usually not too hard to determine by simply listening to the noise at various locations being considered. It is a good idea, however, to excite all of the possible noise sources to determine the extent of the problem. One source that is often overlooked is water flow noise. This is a concern since it is extremely difficult to get rid of this noise. Have someone run the shower while you listen at the proposed location to see if this will be a potential issue. In my theater, this is a problem which I have never completely solved. The water lines run right through the ceiling of the theater and no amount of noise control has eliminated the noise altogether. While not a problem, there are times when it is clearly audible. Again, don't underestimate the potential for this noise problem nor the difficulty of solving it. One is hardly going to reroute the plumbing for the HT.

Appliances then become the next major offender and often the same appliances that cause water flow noise also cause machinery noise, exacerbating the problem. As I discussed in Chapter 6, HVAC systems present particularly significant problems for the HT design. The greater the distance from the HT to these noise sources the better. Although one of my theaters shared a wall with the HVAC room, and I was surprising pleased at the noise reduction, so it is possible to isolate these noises. As I described earlier, the real problem with HVAC is the duct noise. The ducts go throughout the house and, from a noise standpoint, link every room to every other room. Make sure when selecting a location that there will be room to put in mufflers because, as I said, without mufflers there is little point in any further noise control measures.

Of course, each situation is unique and I have not seen them all. However, it is a rare home that does not have some form of HVAC. Most of us simply live in locations that have temperature extremes that fall outside of the desirable range. In the North, the heating system runs much of the time in the winter, and, in the South, the air conditioner runs most of the time in the summer. Few of us enjoy a location that has a stable enough temperature that no HVAC system is required. And remember that we all want fresh air. Opening a window is a sure potential for noise, depending on where you live.

Assessing the noise situation and allowing for room to solve the noise problems is a crucial task that must be performed.

Finally, there are the equipment considerations. It is best to have the equipment placed in a wall such that it is accessible from the rear. This is because of the noise issues and wiring convenience. Many pieces of electrical equipment have fans which tend to be noisy. Noise abatement tasks for the equipment tend to reduce the functionality of the internal ventilation systems making marginal designs ineffective, resulting in equipment reliability issues. Placing the equipment outside of the HT room keeps the noise out. The projector can be arranged to show through a window in a manner completely analogous to a commercial theater. Noise issues are the same reason that there are “projection booths”. I remember the first HT that I every saw was literally a 35 mm projector system. The noise abatement enclosure was enormous with sound proof doors, etc. Solutions to this problem today are much more reasonable.

The next thing to think about are the walls and floor. If the floor of the HT is not concrete, then sound isolation might be all but impossible. At the very least, a floating floor will be required which will take at least 2" to achieve. If there are living spaces above the HT, then an isolated ceiling will also be required. The floor and the ceiling are the two hardest surfaces to sound isolate because of their orientation to gravity. But side walls can also be a major problem. Sound isolating walls must not be attached to a weak portion of the structure or their effectiveness will be seriously impaired. In most HTs that I have been involved with, the walls that were not made of pored concrete were anchored to steel beams. Think of a solid steel support structure as being a mechanical “ground”. If the walls are not attached to ground, then they will excite the rest of the structure by their physical motion, thus propagating noise throughout the structure.

In the example design that I will show later, the selected area was directly beneath the master bedroom. Without adequate sound isolation this HT would be unusable after one of the tenants had gone to bed. The same is true in my theater, although it is not under a bedroom. But without good sound isolation to the sleep areas, a HT can be a major disturbance.

I look for several things when selecting a location. First, I look for a corner which has three concrete surfaces. This reduces the noise isolation problem basically in half. As I will show, these concrete portions are not at all a problem, they are an asset. If this location has windows, I consider sealing them up. I did this in my HT. Then, I look for a major structural member to locate the other walls, or as many of them as possible, again simplifying the sound isolation task. Creating a large distance from the HVAC components is desirable, but I have found it easier to isolate these components than to isolate the HT from the rest of the house, so the HVAC components are not the primary concern. As I said, one of my HT shared a common wall with the HVAC room and this turned out to be a solvable problem. Locating the door—remember that?—is an important consideration because seats, loudspeakers and equipment cannot go too close to the doors without creating problems with ingress and regress.

10.2 Typical Design Examples

I'll start by giving some specific design examples to typify the process. The first is my own home and the second is a client's.

10.2.a Example One

Figure 10-1 shows a rough layout of my home's basement. To get an idea of scale, the length is about 60'. The HVAC components are shown and the heavy solid lines are steel floor beams with the poles shown as squares. Poles are important considerations because they cannot be moved and a pole must not be in the line of the projector or the audience. The region marked "low clearance" is not out of the question for usage, because head-

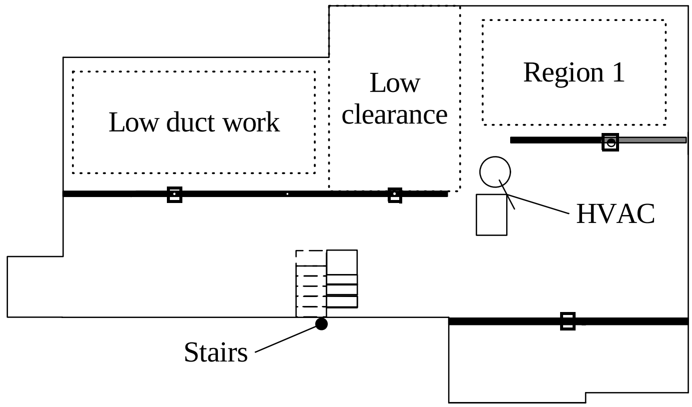


Figure 10-1. Floor plan for design Example One.

room is not a big consideration in a HT. People are mostly sitting not standing, good headroom always makes the project easier.

For a 10' wide screen at 1024 x 768 resolution (the only option for aspect ratio when I bought my projector), Table 9-1 on pg. 159 shows me that I would need at least 20' for this screen. This number comes from 15' for the viewing distance, 3' for behind the screen (for the speakers) and 2' for behind the audience (a minimum). In actuality, I do not recommend placing the viewing location with a wall behind the viewer. That's because of the sound effect of being at a wall. The lack of sound waves impinging from the rear causes a reduction in the perceived spaciousness of the room. I was really looking for about 25' depth by at least 10' wide. Not a small room but not outrageous either. The basement had a few glass block windows but nothing major, so I could ignore the window locations. One of these windows did end up behind the screen location and I simply blacked it out with paint. Without doing this, the screen was completely washed out at sunset as the sun came through this window.

A quick survey of the layout will show that this room would have to lie along the longest dimension of the floor plan and that the only two options were "region 1" or "low duct work". The duct work situation (ventilation for the main floor) caused me to plan around "region 1"—at first. After some initial design work, and for complex reasons, some of which are not germane to this text, this region did not work out. I was forced into "low duct work". Not happy about the situation, I bit the bullet and rerouted all of the duct work in this area either along the walls, behind the screen, or up into the joists. This was a major task which had to be done before I could even start the rest of the project. These kinds of major tasks can come up in this kind of project.

Figure 10-2 shows the detailed layout, along with the dimensions, for the selected site. Some consideration was given for putting the door on the

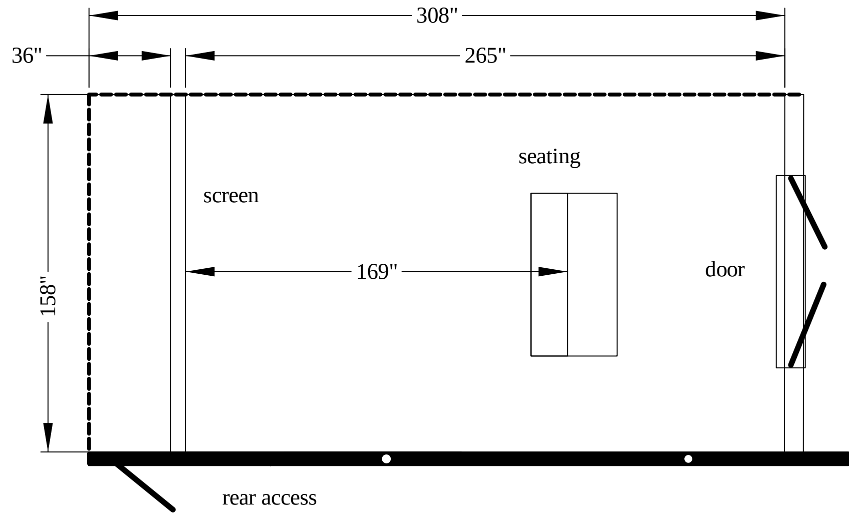


Figure 10-2. Selected HT location showing screen placement and seating location.

side. This has the advantage of freeing up the back of the room for the equipment. The choice shown in the figure forced the projector to be placed in the room as opposed to outside of it. In hindsight, I think that I would reconsider this choice of door location as a sidewall door would have allowed me to move the projector back behind the rear wall. The projector noise remains as the single biggest problem that I have to tend with. My original plan was to mount the projector in a “quiet box”, but multiple failures of the projector convinced me that this was not a good idea. I have to live with the projector noise.

The screen wall is not really a wall at all, it is simply a stretched screen and some cloth covering a wood frame so as to be acoustically transparent. The speakers are in the space behind the screen. The door on the side is necessary only if one cannot move the screen so as to get behind the screen to place and/or work on the three speakers and a subwoofer. My clients room would not allow for this door and he must remove his screen to get in the back. At 14' back from the screen, the seating placement is ideal for a 4:3 10' wide screen.

10.2.b Example Two

As another example of how to place the HT, consider the floor plan shown in Figure 10-3. This basement had a large ceiling height and the HT

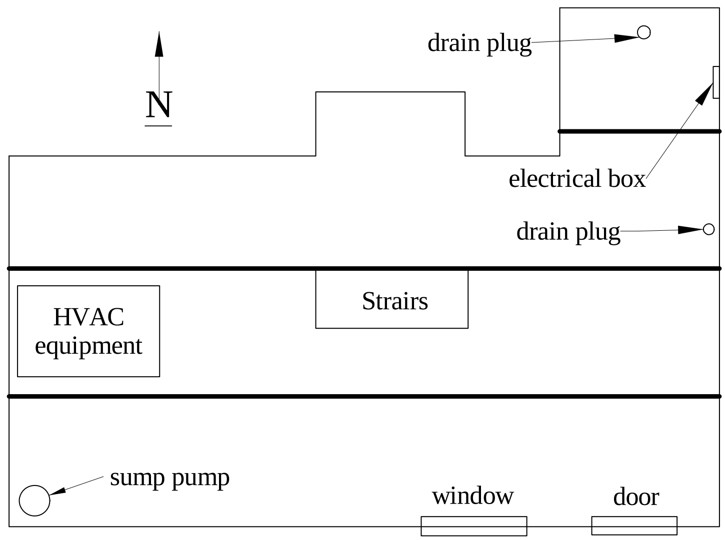


Figure 10-3. Basement floor plan for example installation two.

could have been placed almost anywhere, although there were some locations that looked better than others. I suggested keeping away from the widows and exterior door since doing so would solve a lot of ambient light problems. This eliminated the southeast corner. It would also be nice to avoid the HVAC equipment indicating that the eastern part of the floor plan would be the most desirable. The client had already suggested the northeast corner, and short of a few drawbacks, it was ideal. The major drawback was that it was directly below the master bedroom (which is not evident in the drawing). This made the floor/ceiling sound isolation design (the most difficult part) extremely important.

There were also several features that could not be moved and would have to be designed around. The electrical panel, which is not excessively large, would require that it be hidden and, yet, still remain accessible. There

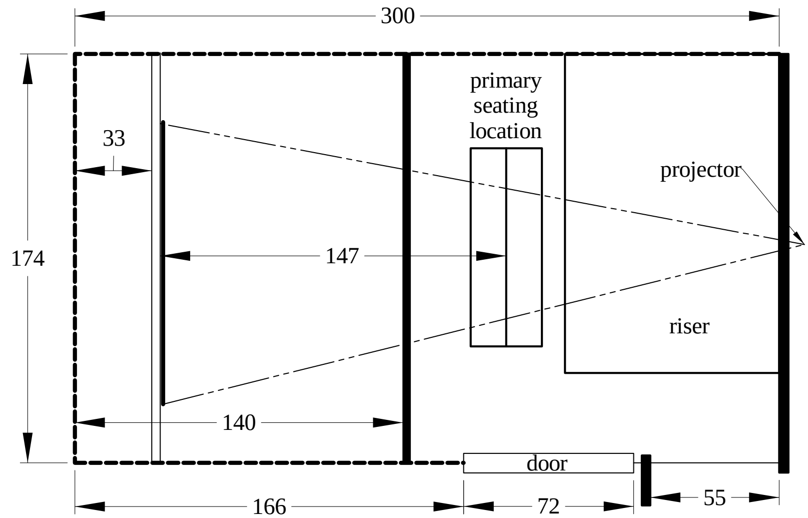


Figure 10-4.
HT basic layout for
Example Two.

were also two drain plugs in the floor that could not be moved. These are all shown in Figure 10-3

Figure 10-4 shows the selected site and the basic floor plan layout for the HT. The beam running through the center of the room and the one at the end were a blessing since this would give us a "virtual mechanical ground" for the ceiling, which, as I said, posed a significant design problem. The client wanted a riser for a second row of seating. I do not usually use risers, but they can be useful. This one became a large array of Helmholtz resonators (see Figure 6-8 on pg. 116). The proposed primary seating location is also shown.

Figure 10-5 shows a photograph of the actual site selected for the HT example that will be detailed in this text. Note the patch of Visquien taped to the floor to test for moisture. High moisture content in the floor dictates certain floor treatments. Note the unmovable features of the water meter and the electrical box, both of which had to be designed around. This is the room of Figure 10-4 looking from a location behind the projector.

The projector would be placed outside of the HT room along with all of the electrical equipment. The screen has an aspect ratio of 16:9 with 1280x720 pixel resolution and will be about 10' wide. From Table 9-1 on pg. 159, we can see that the ideal seating distance for this configuration

Figure 10-5. Photo of the space selected for the home theater system looking towards the location of the screen.

should be about 14'. The seating distance shown above is slightly short of this value. Given this situation, it would have been desirable to have reduced the screen size slightly except that the projector did not have the ability to do this (not enough zoom range). In the end, this situation did not turn out to be a big problem in the final showing, but the seating is a little close for my taste.

The headroom requirements for this project would have been a problem except for the extended ceiling height of this basement—10'. This allowed for the projector to be placed very high, just clearing the viewers heads when seated. This theatre layout would not have been possible in a normal ceiling height of 8'. It is always a good idea to check on all of the projection lines to make sure that the desired screen size can be obtained and that there will not be any interference with the projection beam. Some modern projectors allow for extreme off-to-the-side projection which does have its advantages in very adverse circumstances. Designing for a normal projector, however, always gives one a greater choice of models and prices.

In both of the examples that I have shown, the ceiling posed the greatest sound isolation problem. Both rooms had at least two walls of poured concrete and a concrete floor aiding in the sound isolation through the structure. Example One had a substantial wall are that needed to be isolated but, fortunately, this wall could be attached to a steel beam, thereby easing the design considerably. Both rooms would need substantial mufflers for the HVAC connections into and out of the rooms since there were adjoining living spaces that would be sharing the HVAC ducts. In both cases, significant attention would have to be paid to acoustically de-coupling the HVAC system components from each other.

Neither of the two examples that I have shown had to consider light pollution, except through glass in the doors, which is easily accommodated with simple shades, since the light is not directly impinging on the doors in either case. Both rooms had high output lights just outside of the HT door which increased the ambient light conditions in the HT, if uncontrolled. But in both cases, these lights could be turned off or reduced, which is virtually always required in practice for both of these spaces.

10.3 Considerations for Locating Equipment

In Example One, there was no provision made to place the AV equipment outside of the HT room. There were several reasons for this, but they are not germane here (I can't remember them now!) The fact is that noise from these components became a serious problem and remains a serious flaw in this project even today. As I have said over and over, never underestimate the potential for a noise problem to muck-up things.

No matter how low in noise a manufacturer claims his projector is, it is probably too noisy to be in the HT itself without some form of noise abatement. This is not an easy task because the noise is virtually always generated by cooling fans. Without adequate cooling the equipment failure rate will skyrocket as I found out after enclosing a projector in a sound box. Even with forced air cooling of the box itself, it still overheated and shut down. The units cooling capacity was marginal to begin with and it went over the edge with even the slightest restriction in the air flow.

The thing to keep in mind is that smaller is noisier. That's because smaller cases need smaller fans moving air at a higher rate. The fan has to turn faster, causing more mechanical noise and the flow noise goes up as the air velocity increases. The one thing that one never wants to do in a HT is to buy a projector because it is small. This only ends up being a big problem.

In all of my designs, a computer plays the central role of control for both the audio and the video, I will talk more about this in later chapters. Computers are always noisy devices, once again it is the cooling fans that are the culprit. Usually, it is the small fan that is placed on the CPU. A rule of thumb is that smaller fans make more noise. If you want to reduce the noise level, increasing the size of the fan will usually work. The speed can be lowered, which also lowers the air velocity and flow noise, so you get a double benefit. But, be careful because not all fans are created equal. Look for a fan that turns smoothly when you rotate it with your finger. If it jerks, or cogs, then this cogging causes a vibration which is conducted through the mounting to the motherboard and the case. This vibration increases the noise level radiated from the case and sometimes causes rattling, which is most annoying. Enclosing the computer is virtually always a necessity and moving it, along with the projector, completely out of the room is the best solution. In Example Two, this was the approach that was taken and with a sealed door on equipment cabinet, the computer noise was non-existent. The power amplifiers also had fan noise and this sealed (from the HT) equipment cabinet proved to be a very effective solution to these problems.

Planning ahead for wiring is also a good idea. Some wires are easy to accommodate, like speaker wires etc. but others are not. The video signal wires are notable in this regard. These cables carry very high bandwidth signals and cannot be long without very expense line amplifiers. It is a good idea to consider the locations of the video processing equipment and the projector early on in the design. Keeping them close together is the best bet, because large separations will require some specialized solutions.

10.4 Conclusion

The initial planning for the HT is a crucial stage that can easily make or break the project. Proper choice of location is the first decision. You may want to consider several different locations and examine the pros and cons of each. The location may change after due consideration. Be conscious of the ambient condition of noise and light as well as the locations of spaces that could take offense to a HT. Consider the equipment location and wiring carefully.

CHAPTER 11 CONSTRUCTION

In this chapter I will talk about actually constructing the room using the knowledge that we have gained from the rest of this text. It would be ridiculous to think that I could tell you all that you would need to know to actually construct a room from scratch. There is just too much material to cover that entire subject. What I will discuss are those construction techniques that are unique to a HT design and construction and are not common in the construction trade.

Readers will generally fall into two categories; 1) those with considerable construction experience who might be doing the work themselves or 2) those with little or no experience who will (hopefully) be hiring someone else to do the work. If you do not fall in one of these two categories, then you will have to do some extensive reading elsewhere to learn about using the equipment and how to perform the actual construction techniques that are used in the building trade. You may even find that you will need to learn and understand the building codes. If you do plan to do the work yourself you will need to have some extensive specialized equipment, like a table saw, to complete the work. One cannot even consider doing this job without a substantial investment in construction equipment.

If you're hiring someone else to do the work and they are not familiar with the construction techniques that I propose here, then you will either have to teach them or they will have to read this chapter. As I have said over and over and I'll say again, many construction techniques commonly used in the trade do not take sound isolation into consideration and are extremely poor in that regard. Building a HT without sound isolation from the rest of the structure is not recommended unless you live alone and well separated from other people. That is the only way that I can see building a non-isolated HT. And lest we not forget, to get the proper acoustics in the room these same basic techniques are still recommended. It is not just for sound isolation that these techniques are used but for the rooms proper interior acoustics as well.

11.1 The Foundation

I won't say much about the foundation of the HT because if the foundation is anything but a concrete slab, then creating a suitable foundation for the HT is a project in itself. Isolating the foundation of a HT from the remaining structure in the non-slab case can be done, but the techniques are well beyond the difficulty level that I will be showing in this chapter. If you are unfortunate enough to have the HT on a common suspended foundation, such as wood floor joists that are supported by steel beams, then floor isolation poses a major design and installation problem. Basically, one has to float the entire room on a resilient base of foam or some other soft material—not an easy task. There will also be no mechanical "grounds" (see pg. 186) that I utilize in my designs.

The techniques and materials that are used for floor isolation are highly dependent on the specific situation and it would be difficult to generalize a technique that was universal. This means that each case has to be examined by a knowledgeable person in order to determine what can be done and how to do it. I feel qualified to make this design assessment, but I am not confident that I could relay this knowledge in a way that would be meaningful. If you will need sound isolation of the floor from the rest of the structure, then you should seek professional advice.

11.2 The Walls

Wall construction is quite common and I am sure that everyone has seen standard 2x4 wall construction with studs 16" on center. I will not show a drawing of this configuration as it is so common that it would not be useful to showing it. However, I will describe some of its drawbacks.

Depending on the building codes in effect in the specific area, the wall may be made with 5/8" or 1/2" drywall or Gypsum board attached to the studs. This wall board material is actually quite light—low density—which is of course why it is so attractive, it is relatively easy to handle. But this low density also means high sound transmission (low TL) as shown in Figure 6-6 on pg. 112. Note that the transmission loss in the mid frequencies is controlled by the mass—i.e. it is mass controlled. A low mass density, and, hence, a low total mass will mean low values of sound transmission loss for these panels in this frequency range. Simply stated—sound goes right through normal wall material. I think most of us have already come to know this.

To make matters worse, the fact that the wall has a common structure between the two sides means that sound waves impinging on one side of the wall will physically move the entire wall re-radiating the sound on the other side. This is secondary to the sound that simply passes through the drywall, and resonances of the wall structure can actually cause the sound levels to increase through this mode of transmission. At low frequencies, this mode of transmission is the dominant one.

The difficult question then is how do we make a wall that does not pass sound? Two drawbacks of existing walls must be corrected. First, and foremost, we must make the wall heavier, which is the easier part, but second we must de-couple the inner and outer surfaces—the two sides of the wall. Then, we must consider the internal damping of the wall as discussed in Chapter 5 (see pg. 97).

CLD
Constrained layer damping—a construction technique of placing a well damped mastic between two stiff panels.

There is a technique that I will discuss for making the walls that deserves some background explanation. It is know in the noise control business as **Constrained Layer Damping**.(CLD). This technique has been widely used in automotive industry as a way to dampen out the highly resonant sheet metal panels that make up a car. No doubt everyone has seen the typical flexible, heavy mastic glue that is commonly used on sheet metal panels in cars to dampen these panels.

Figure 11-1 shows this situation on the left hand side of the figure. When the panel bends, as shown in the figure, the mastic layer is carried along being compressed on the inside edge and expanded along the outside edge as shown in the enlargement of an element of this material. There is virtually no shearing of the element as the panel bends. The material in this configuration has a predominately added mass effect and yields only a small amount of damping since there is only a small amount of internal motion of the mastic itself.

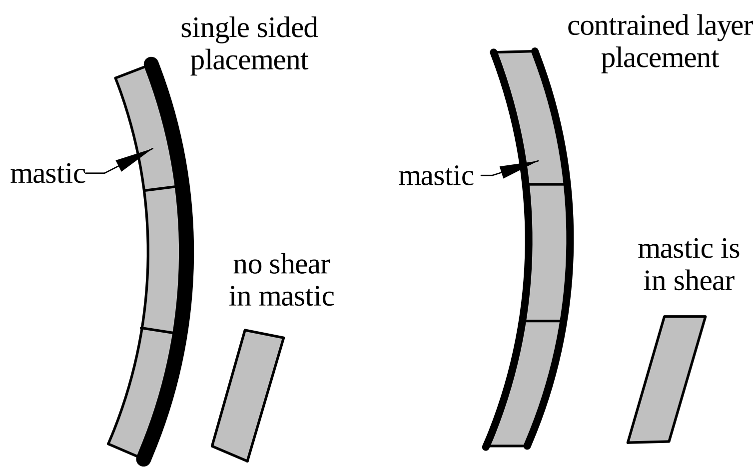


Figure 11-1. Example of a panel with damping layer constrained and unconstrained.

On the right hand side of the figure, I have shown two panels, each half as thick, but now the mastic is place between the panels. When this panel configuration bends—it has about the same bending stiffness as in the previous example—it places the mastic into a significant shear displacement as shown in the blowup of an element in the lower right hand corner of the figure. The internal motion of the mastic is greatly enhanced in this configuration. It is this internal motion of the mastic layer that dissipates the vibrational energy as heat due to the friction between neighboring particles of the material. I will refer to this technique throughout this chapter as CLD.

In practice, the mastic layer in a CLD configuration can be made much thinner than that used in the unconstrained layer situation (on the left in

Figure 11-1) and yet achieve an order of magnitude greater damping effect on the panel itself. This principle is an extremely useful one to know and it is one that I use in almost every situation where there is a panel that I wish to dampen. Rather than simply placing some compound on the panel, I spread a thin layer of mastic and then place another thin, but stiff, panel over this mastic layer. Sometimes the second panel will need to held in place while the damping material sets up.

It is easy to see how this technique can be used in a wall structure to accomplish two of our goals. The first goal is sound isolation. The CLD panel does not transmit sound as effectively by converting it to energy. Second, it adds considerable low frequency sound absorption. This second attribute may not be obvious, so let me explain. By making the surface of the CLD wall somewhat hard, like sheet rock painted with enamel, its reflectivity will actually be rather high whenever the panel is rigid—does not move. Since the panel is basically a limp suspended mass, it will move less and less as the frequency goes up, becoming virtually rigid at mid to high frequencies. With a fairly hard surface treatment it will basically act as a damper at lower frequencies, one whose damping coefficient will decrease with frequency—which is exactly what we want! I cannot over state how important this feature is, and I know of no another technique which has this feature.

I do need to point out that this panel construction must be free to move for the CLD effect to work. The most effective means for doing this is to mount the panels on springs. Spring mounting the drywall is actually a fairly common technique for wall construction where sound isolation is desired (common technique, seldom used). There are even specific mounting materials made for this purpose. In particular, I refer to RC-1 Resilient Channel made by USG, among others (under different brand names). It is basically a thin piece of strip steel folded along its length and constructed to act as a spring between the studs and the wall panels. The assembly technique is to mount the resilient channel horizontally to vertical studs and then to attach (hang) the sheet rock panels to the channel with dry wall screws. The panels then become basically freely suspended and if properly done will have a low resonance frequency. It should be pointed out that with RC-1 channel and a single sheet of drywall the system has virtually no internal damping. Only through CLD does this construction technique provide any appreciable amount of internal damping.

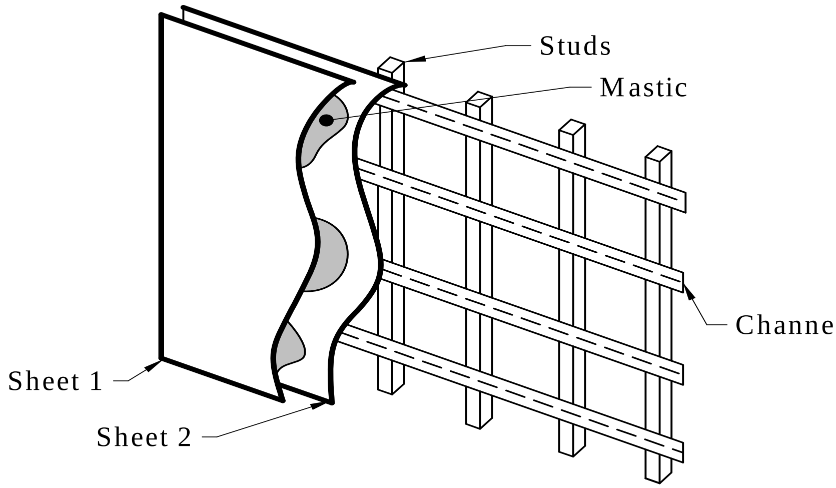


Figure 11-2. Schematic drawing of wall construction showing CLD technique on resilient channel.

Figure 11-2 shows a simplified drawing of the manner in which a resilient panel CLD wall is created. Starting with a basic studded wall, the channel is attached to the studs horizontally and the first sheet is attached to the channel. Care should be taken not to (accidentally) attach the panel directly to the studs by screwing the dry wall screws through the channel and into the studs. This will defeat the entire construction. Next, a second sheet of dry wall is attached to the first with a damped mastic layer spread uniformly (reasonably uniform) across the back of the second sheet. The second sheet is usually supported with dry wall screws to hold the two sheets together until the glue dries. For best results, these screws are then removed, but this is not an absolute requirement. The problem gets be to down-the-road when the screws want to back out—not actually being anchored to anything. The glue layer (mastic) should not dry to be rigid but should remain flexible. A non-hardening glue is required, but it must also be one that will not continue to flow—like a flooring glue, which would tend to let the second sheet fall off eventually. I have used *Liquid Nails for Sub-floors* which works well, but there are many other options. Silicon or latex chalk could also be used. These don't dry as hard as the Liquid Nails, but they also aren't as well damped. There is no reason why a mixture of the two adhesives would not work, although I don't mean chemically mixing them. I don't know what that would do. Using different areas of each was what I had in mind.

Figure 11-3 shows a photograph of a CLD wall construction shown from the rear. This photo highlights several features. First, there are stagged studs in this wall (which I will talk about in a moment), and, secondly, the technique used to conceal the beam support posts, which are always present, is also shown. They are embedded in the wall. The HT door can just be seen on the left. In this photograph, only every other stud is attached to the channel. The other studs are used as support members for the wall to be placed on the near side of this wall. The wall surface for the near side has not yet been installed. The studs and the channel are both spaced at 16" on center.

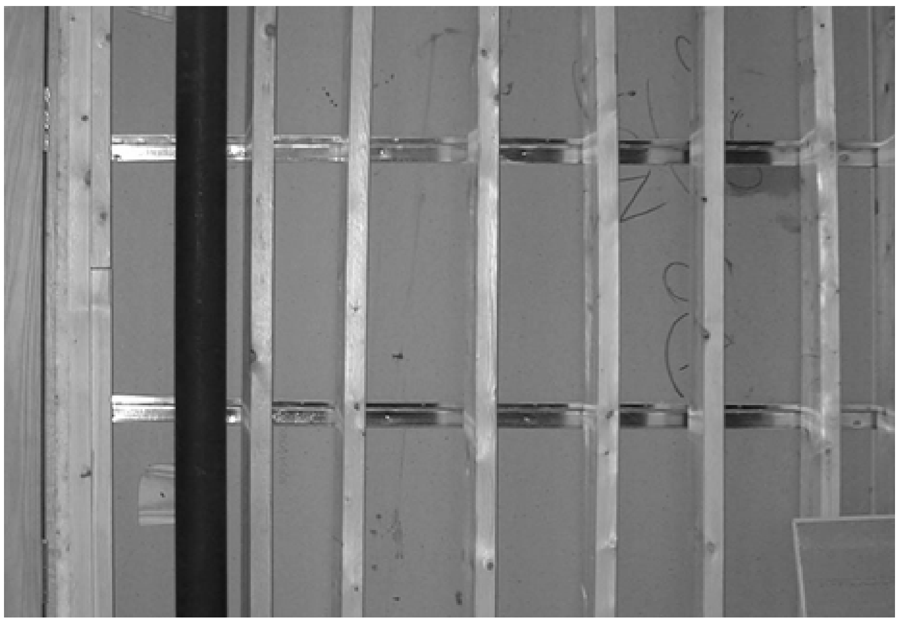


Figure 11-3. Photo of an actual installation of a CLD wall shown from the rear. This is a double studded wall with a door jam also shown.

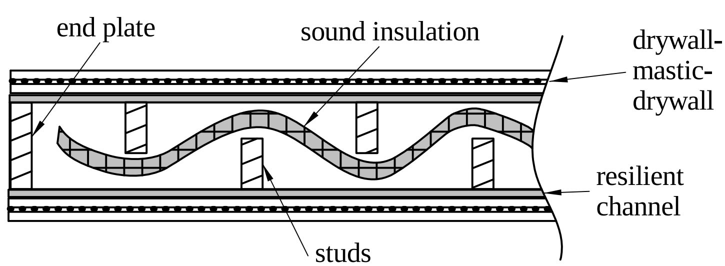
As I have mentioned before, the wall in our HT needs to serve two purposes. A low frequency damping function is performed by the CLD as described above, but this technique is not usually sufficient in and of itself to yield the kind of sound isolation that we want. The low frequencies in a HT can be particularly problematic to isolate because of explosions and such which are mixed on the film at very high levels for impact. These low frequencies will shake every part of the room and this shaking needs to be isolated from the remaining parts of the structure or it will re-radiate as sound. This is called **flanking transmission** and it can be very difficult to control. It basically means that every single structure inside of the room

Flanking transmission the sound and vibration that goes around our noise control measures.

must not be touching any part of the home external to these interior components. This is actually impossible since both the home and the HT must both sit on a common foundation. The foundation difficulties have already been discussed, but we can see here that even the connection of the vertical walls to the floor can become a real problem when foundation isolation is required.

Returning now to the isolation problem for the walls, there actually is a technique in use which works quite well. The idea is to have staggered wall studs such that each wall has its own independent set of studs. Figure 11-4 shows an example of this double stud wall construction. The top and base plates for this construction are standard 2"x6"s, as are the end pieces. The interior studs of this wall are staggered so that every other one touches only one side of the wall, while the other set touches the other side. The studs are spaced on 8" centers, which ends up being 16" on center for studs along each wall surface. Sound absorption material is usually draped between the studs for extra sound absorption of the sound that does get through this wall construction. The sound attenuation of this configuration can be quite large, up to about 50dB TL. That is if it is properly attached to the structure.

Figure 11-4. Schematic of staggered stud wall construction. (top view)



It is important to attach the above construction to rigid points in the surrounding structure. Basically, a rigid point is a virtual vibrational energy ground. This point has to be able to be subjected to large amounts of vibrational energy and not move—analogous to current into an electrical ground. This is why I always look for steel beams and concrete portions of the structure since these are grounds which can be used to anchor the floating vibration isolation components. This is why the steel beams were highlighted in Figures 10-1 and 10-3. These beams and the concrete walls are vibration grounds and should be used as the periphery of the HT whenever possible. This was done in both of the examples in Chapter 10. The double wall construction is glued and/or press-fitted between the beam and the floor, basically “grounding it”. Care must be taken with walls that cannot be grounded. They must not be rigidly attached to the surrounding structure at any point that is structurally weak, like floor joists, etc. This will basically “short-circuit” the vibrational energy and create a large flanking transmission.

In the layout of Figure 10-4, the only portion of wall which is not grounded along it length is the 55” section to the right of the door (I’ll talk about the door in a moment). This section was “grounded” to the beam posts at each end and the center of this section was left floating.

When a section of wall is left unattached to the building structure, it is good practice to glue it with some non-hardening glue, like *Liquid Nails for Sub-Floors*. I have also found that urethane foam can also work as a good glue. Most people find this rather surprising, but this foam, when setup, is actually quite strong and it has a structure that does not transmit vibration. In fact, I use this material by the can-full, typically buying a case for a HT project. Any space that has the potential for being a sound leak is filled. When setup the foam trims nicely and seals out sound and vibration extremely well. It can actually be “mudded” up and painted, thereby becoming part of the wall.

Figure 11-5 on pg. 206 shows this foam being used in several areas around the door. In the next section, I will describe how this foam is actually used to support the door, becoming its only attachment to the surrounding structure. I once referred to my HT as being put together with urethane for its usage was so predominate. I have been know to fill spaces as large as 6" by 12" with foam by creating a supporting grid of paper or some wire mesh and filling the area with foam. When set up this material works just fine in this application, but it is rather free flowing until it sets up and must be supported in some manner to prevent it from flowing out of the space that it is placed into. It can be trimmed, grouted, smoothed and painted and it becomes indistinguishable from the wall. When partially setup, the material can be modeled into shapes, although care should always be taken not to get it on anything of value because it is difficult to get off after it sets.

The procedure for partitioning the HT room from the rest of the house is to find good ground points, attach double studded walls to these points and

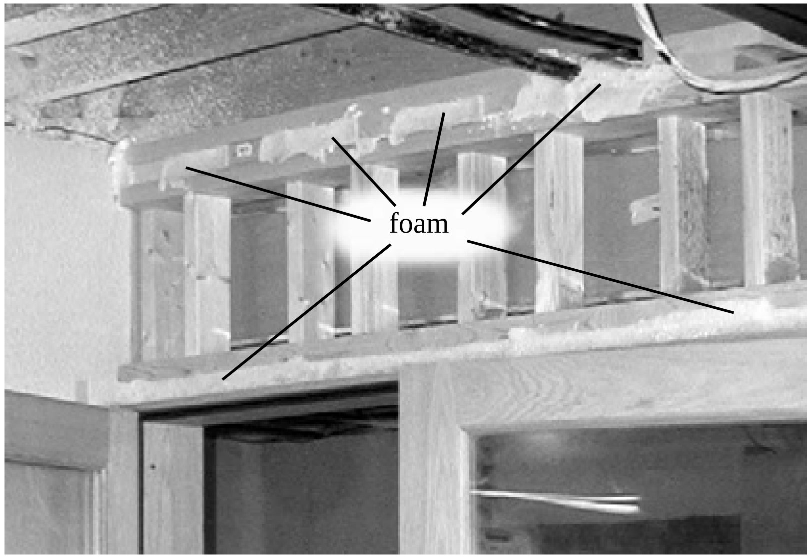


Figure 11-5. Photo showing the use of Polyurethane foam for sealing and attaching structural components of a HT.

fill in all gaps with urethane foam. This will create the basic space which is to become the HT. The next step is to put in the door and finally to install the all important ceiling.

11.3 The Door

Of course, the room must have a door, and some, like mine, will have two. The first door is the entrance for the participants, and the second, if needed, is for access behind the screen. The screen access needs to be made by an external door unless the screen is removable. Having done both, I strongly recommend making the screen removable since this is far easier to do and works just as well. In one case it simply swings up and out of the way—very convenient.

The main door is usually a double (dual) door. This makes for a nice appearance to the theater, a more inviting entrance way than a single door. It does not have to have glass in it although I have found this to be aesthetically pleasing and not really that much more expensive. If there is a light pollution problem in the space just outside the door, then the window may not be such a good idea. In either case, the door itself is important and warrants some discussion.

Interior doors just do not fit the bill. If not properly sealed, the door will be a short circuit to everything else that was done to isolate the sound inside the HT. Fortunately, the requirements for a HT door are identical to those for an exterior door. These exterior doors are heavy, use double pane glass and always seal well around their periphery. They really are ideal, although they are certainly more expensive than interior doors.

Usually a custom door has to be ordered, which is not really a problem since there are an numerous companies who do this kind of work. In fact, these doors can usually be ordered with the extra wide jams that are required for the double studded walls, which are much thicker than normal walls.

The choice of door is of course, a matter of taste, but I prefer solid Oak doors. They are readily available, can be obtained with solid oak jams and threshold and can be purchased pre-hung, which is a real advantage to the less capable carpenter. In every case of HT that I have done, this is the door that was used. Once the client sees the appearance of the double French door and what a spectacular entrance-way it makes, it has been their unanimous choice. These doors arrive pre-assembled and once brought into the correct position are easy to install. One does have to be careful to insure that they can actually get the door into the space since they can be quite big and were never intended to be moved to the interior of a home. In my second HT, I had to disassemble the door to get it down the stairs in pieces, and then reassemble it in place. This was still easier than trying to hang a door like this myself.

Once the door is in place, I recommend gluing it there with urethane foam. This actually works quite well and fears of this mechanism not being strong enough have never been justified by anything that has occurred in practice. The only problem that one has to consider is that the foam does expand on setting up and it can do this with a great deal of force. The door must be held in place with screws, which are then removed later, after the foam has setup. The door is placed inside of the opening in the room’s wall structure and should have a gap of at least 1/2" all the way around. I recommend actually using a 1"–2" gap. The foam has no trouble closing up this gap and the larger the gap the less vibration that will be conducted to the walls by the door, and vice versa. Remember that the door is rigid, unlike the walls, and will move due to sound impinging on it. If this vibration is allowed to be conducted to the walls then it could be re-radiated as sound. The foam mounting prevents this.

As I said, the door must be held firmly in place while the foam sets up, because the foam can and will cause the frame to bend and bow if it is not. Then, when the foam sets up, the door will be out of alignment, which can be a real problem. I learned this lesson the hard way. Don’t underestimate the strength of the urethane foam once it sets up. By using screws or some other method of support, the foam’s expansion won’t warp the door and it will open and close freely. It should be noted that there should not be any connection between the door and the supporting structure that cannot be removed. Excluding, of course, the base which I have assumed in on a concrete slab and is grounded.

The area all the way around the door frame is then completely filled with foam. To avoid getting the foam on any nearby parts of the room that the foam might damage, you may want to mask these off with tape and paper, much like you would for painting. After the foam has set up it can be trimmed and the doors supporting screws can be removed. Some screw holes may have to be filled in. The door should now be rigid and open and close freely. The foam around the door can be clearly seen in Figure 11-5.

For the access door, which is not visible from the HT, I typically just use an exterior grade steel door because they seal well and isolate the sound transmission quite effectively.

11.4 The Equipment Rack.

A HT will have a substantial amount of electronic equipment associated with it and some of it will make noise. Therefore it is advisable to provide a cabinet to contain this equipment. These can be of two forms, namely, interior and through the wall.

The interior form of cabinet is basically a stand alone enclosure box which has a sealed door on the front and possibly the back (see Figure 12-8 on pg. 242). Some consideration must be given to ventilation in this configuration. In this case I built the enclosure hinged on one side and attached to the wall, so that it could swing out to expose the wiring at the rear of the components. There was a Plexi-glass framed door on the front which sealed all the way around. The enclosure was open on one side of the rear for ventilation and the top was partially open for the same purpose. The noise from the components does radiate out one back edge that was open, but since this side faced away from the audience this noise was not objectionable, although, in low ambient conditions, it is perceivable. I have found that this form of enclosure is not really the ideal, and I now always recommend using a through the wall design, even thought they too have their downside.

The through the wall enclosure is basically a box with an open or partially open back that is placed into one of the walls of the HT. This box is supported on the floor (vibrational ground) by a rigid structure and is set in the wall in a manner such that it does not touch the wall. The enclosure is then sealed in place using a technique identical to the door—urethane foam all the way around the enclosure. A glass frame door which has air tight seals is then placed on the front. This method has proven to be extremely effective at noise abatement.

Figure 11-6 on pg. 210 shows a HT with the opening for the equipment cabinet on the right hand side of this photo. There is a supporting structure at the bottom of this opening, outside of the room, on which the equipment cabinet will be placed. It was professionally made by a kitchen cabinet maker out of oak with tight fitting doors that would seal.

Figure 11-7 shows this same section of the room, but more nearly complete where the equipment cabinet has been installed and has its doors attached. This form of installation of the equipment has proven to be vastly superior to any of the others that I have tried. It is virtually sound proof—with the door closed—and allows for optical remote control of the equipment inside. For wiring purposes, it is easy to get at the equipment inside from the rear and it also has good access from the room’s interior. This enclosure had a small ventilation fan mounted in the top and a vent hole in the bottom to draw off the heat from the equipment. All in all, this form of enclosure is strongly recommended.

Figure 11-6.
Photo of rear wall of HT under construction showing the opening for the equipment cabinet and the passive absorber riser structure.

Figure 11-7.
Completed back wall of HT showing door, equipment cabinet, riser and oak paneling. Note the sealing glass doors on the cabinet.

11.5 The Ceiling

The dreaded ceiling! By far, the most difficult isolation problem, I almost always select the room layout based on how well it lends itself to a good ceiling construction. For obvious reasons, the ceiling must be floated, or virtually all of the sound isolation will be lost. But floating a structure as heavy as this is not a trial task.

I have used several techniques for ceilings. Each one has worked well, but each one had its pros and cons. I will describe the general approach and follow that up with some specific examples of how I handled the problems in my installations.

Don’t be lulled into thinking that you can use a classical suspended ceiling. While these ceilings are touted as being “acoustical”, they are anything but. A basic suspended ceiling yields a large high frequency sound absorption to the room in which it is placed, but virtually no low frequency absorption. It also offers almost no sound isolation at all. If there are living spaces above the HT, then this will be the most critical surface for sound isolation because it is common to both spaces—unlike the side walls. In one of the HTs described in this text, the master bedroom was directly above the theater. With a sound isolated ceiling, this theater would have been unusable whenever anyone was trying to sleep or using the bedroom for any purpose whatever. Sound isolation of this surface was a high priority.

In one case, I used a combination of classical suspended ceiling tiles and a false ceiling that was itself suspended—supported by the side walls. This was basically a room within a room as shown in the top drawing in Figure 11-8 on pg. 212. The ceiling was constructed of large 4' x 14', the span being 14'. The panels were made up of 2" x 6" outside frames with 2" x 4" stringers inside while 1/2" plywood sheets were glued and screwed to the top. These panels would probably not have been strong enough to span the 14' without bowing, and there was never an issue of load strength. They only had to carry their own weight. But there was a concern that over years they would bow from the dead weight, creating an appearance issue.

The bowing concern was alleviated by placing a 1/16" piece of strip steel 6" wide and 14' long between each panel as it was being installed. By gluing and screwing together the panels with the steel sandwiched in

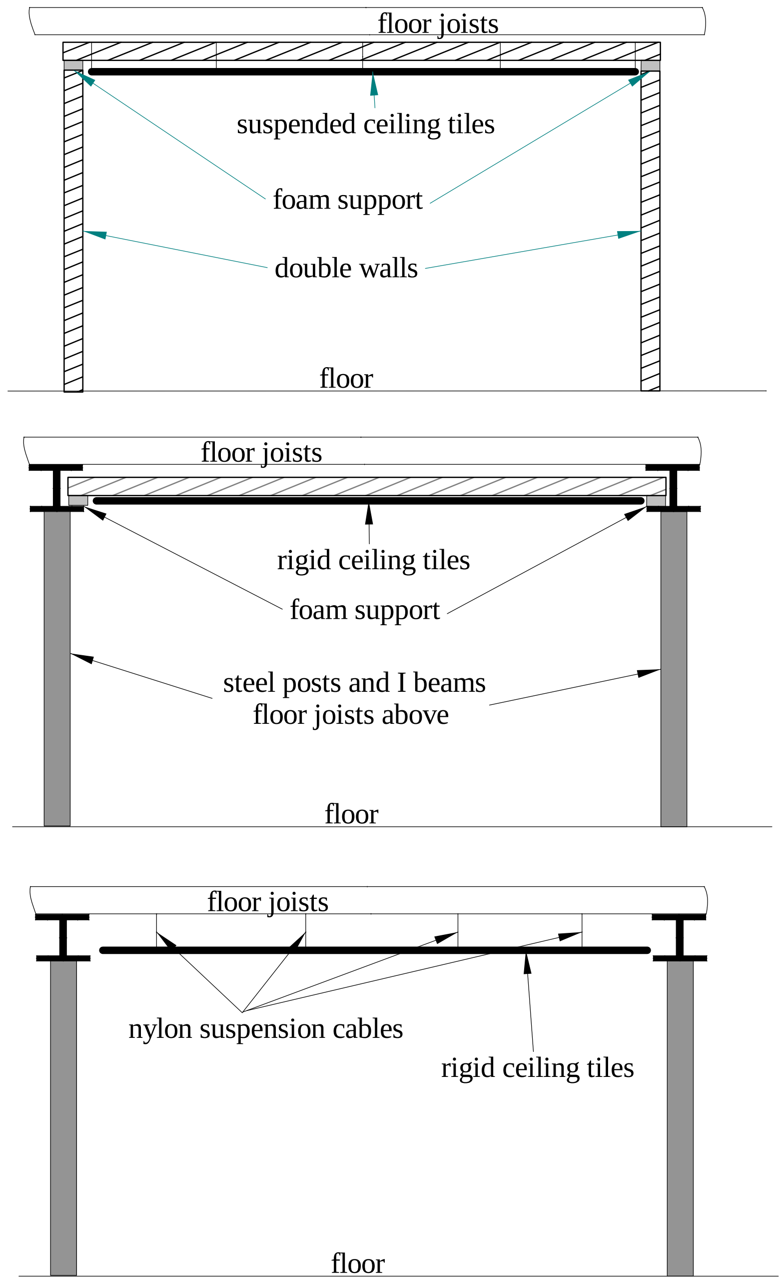


Figure 11-8. Three methods for suspending a ceiling in a HT. Method I - two supporting walls with a false ceiling structure supported across them. Method II - wood beams supported across the steel floor beams with rigid ceiling tiles attached. Method III - rigid ceiling tiles supported from the floor joists via nylon rope.

between, I was able to create a situation which would be extremely resistant to bowing, since the steel would add tremendous strength in this configuration. The room was 14' wide so the panels had to span this distance, which would normally require a 10" or 12" tall joist, without the worry of bowing. Ceiling height was critical and the addition of the strip steel allowed for the use of 2" x 6"s and bowing was not evident even five years later. This construction is shown in Figure 11-9.

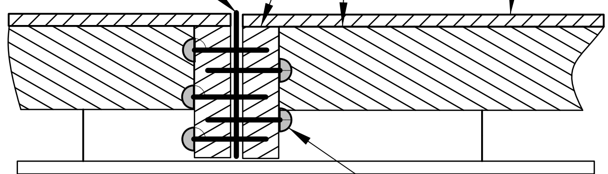


Figure 11-9.
Detail of the suspended ceiling panel attachment construction.

When the ceiling is supported on its periphery, as shown in Method 2 of Figure 11-8, then, it has to be structurally isolated from the supporting walls. This is to prevent the wall vibrations and ceiling vibrations from propagating into one another, i.e. the ceiling vibrations don't excite the walls to reradiate the sound and visa-versa. This is also done with the use of urethane foam. By supporting the ceiling structure on small wood blocks, about 1" tall, the remaining gap can be filled with foam, both as a support and a sound leakage sealant. Then, when the foam has set up, the blocks can be removed and the remaining holes filled with foam. This situation is shown in Figure 11-10 on pg. 214. The foam has been inserted all around the structural member which is temporarily held on a small wood block. The block is then removed and foam is inserted in its place. The final support is strong enough to easily support the weight of the ceiling, yet it is flexible enough to isolate vibrations from going through it. A support that carries a load, like the one shown here, or as another example, an engine mount, is always a critical point of vibration transmission. The high load (due to gravity) always makes the reduction of vibration transmission across supports that act against gravity a difficult task. This is why isolating a floor is a very difficult problem.

*Figure 11-10.
Photo showing method of temporary support for ceiling members and foam for permanent support.*

The structural member shown in Figure 11-10 is but one of a grid of members for a ceiling of the type shown as Method II in Figure 11-8. The entire grid pattern is shown in Figure 11-11. The main 2" x 6" support members rest on foam at each end. The cross braces are 2" x 4"s and are glued and screwed into the main supports. The end results is a grid that is 2' by 2' on which will be glued and screwed 3/4" oak plywood panels. Some of the panels will not be glued to allow for access to the area above the ceiling.

Once the grid structure is up and squared off (an important step) the space above the panels can be loosely filled with sound absorbing material, usually fiberglass. This is not critical but does help with some transmission loss by absorbing what little sound does get through the ceiling, usually at the lights and vents. The ceiling panels are then attached to the grid using a minimum of screws, relying mostly on the glue for support. Figure 11-11 shows the ceiling progressing towards completion. The light installation is evident in this photo.

The light cans must be sealed or they might become a major sound leakage path. Fortunately the number of holes in these cans is minimal and sealing is not difficult, although the method of sealing must be resistant to heat

*Figure 11-11.
Two views of the grid for supporting the ceiling panels. The main support beam (steel) in the center of the room can be seen in the center of the photo. All grid members are supported at one end by a steel beam. One is supported on both ends and one by end supports. The end support at the wall can be seen near the center of the photo (See Figure 10-4).*

*Figure 11-11.
Ceiling under construction showing the grid, the fiberglass fill, and the ceiling panels being installed.*

and the cans themselves must have thermal protection for safety. Usually, the lights in a HT are not on for extended periods of time, so I have never found the heat load to be an issue. In a multipurpose room, where these lights might be on for extended periods, this may be an issue. Another reason to consider a dedicated HT room installation.

I should warn the reader that I have found heat buildup in a HT to be a problem when there is a large group of people in the room. This is because of the extreme amount of sound isolation that these rooms have, which also makes them extremely thermally insulated. It is difficult to get enough air flow through the room to keep it cool when half dozen or more people are in it. This situation is, of course, more seasonal, since the floor and wall will take on a temperature difference with the season. It is also not that common, in my experience, to have this many people in the room at one time, but it has occurred at parties, demos, etc. Usually, there is a lot of flow of people in and out with these larger crowds which helps to circulate the air.

At this point, I have shown how to select the location for the room, how to build the room with good sound isolation from the walls and ceiling and, hopefully, the HVAC system has been looked after according to Section 6.3 - Duct Noise. At this point, the room that I have been documenting looks like Figure 11-12. It is very live, but is well isolated from the rest of the structure and, hence is very quite. Sound from inside of it will not get into the rest of the home and vice-versa. Success at this point will be readily apparent—the room is live yet quiet. If someone is talking or playing music in this room or outside of it, then these sounds should only be barely audible in other spaces.

It usually comes as quite a surprise to most people just how live these rooms can be at this point. In fact, they are unusually live—probably too live. This is because, from this point on, everything that we will do to the room will tend to deaden it. My opinion is that it is far easier to obtain the correct reverberation characteristics when starting from a live room than when starting from a dead one. If one has a dead room at this point, then there is almost no effective way to liven it up. On the other hand, if one has a live room, then there are any number of highly effective, predictable and useful ways to make these rooms less reverberant.

One thing that I have not really covered yet is what to do with standing walls, ones that we have not constructed should they exist, and often do. Of course, these are far easier to deal with if they are rigid. If the standing walls are not rigid, but of a typical construction, then they will not be sound isolated and they could become a weak point in the construction. There are actually any number of techniques that would work in this case. A separate wall could be constructed in front of the standing wall—one that does not touch the original wall. This wall would be made along the same lines as I have discussed in this chapter—but only half of it.

Another consideration for a standing wall is that if it is a rigid wall, like poured concrete, then something should usually be put on this wall for a number of reasons. One is appearance, another is sweating and thermal aspects and there is also an acoustical consideration. My preference is for cultured stone. The rough surface of this material will creates an attractive wall that is well insulated and has desirable acoustical characteristics. A flat wall will reflect an almost perfect image (in an acoustical sense) which is not what we want. A rough wall will reflect the sound waves off of it in a more incoherent manner. The incoherent reflection arriving at the listener

Figure 11-12. Completed room structure shown looking towards the screen area. Note the cultured stone on the two side walls and the frame for the screen made out of PVC. Also note the wood trim along the ceiling panels and the cove lighting at the side walls.

off of this wall will improve the subjective aspects of these early reflections. The reasons behind this phenomena are too complex for this text, but it is widely recognized that an incoherent reflection is preferred over a coherent one.

The details of installing masonry over a concrete wall, or one can actually do this over a constructed wall, are all well documented by the material manufacturers. I won't elaborate on them here other than to say that this is not the easiest job for the novice to perform. I have seen a novice (we're all novices at first), including myself, do an admirable job, it is just hard work, that's all.

11.5.a Modifying An Existing Room

If, as I am sure will happen often, you already have a room that you are going to use, then you need to make a serious assessment of the situation. That evaluation should consider how close this existing room comes to what I have shown you in this chapter. If you are comfortable that it does fit the intent of what I have described here, then great, you are ready to move on to the last and final chapter—the fun stuff! On the other hand, and again I am certain that this will be the more common situation, if the room lacks many, if not all, of the features of the rooms that I have described, then you have to make some serious decisions. Decisions that will have a major impact on the final product.

A room that has poor acoustic isolation, which is true of virtually any normal construction, will have two major drawbacks. First, the sound leakage will cause the reverberation to be low—the sound all leaks out of the room. Second, everyone in the house will be hearing the movie even if they are not watching it. To me, these are both disasters.

Let me give you an example of what I mean. I have a five year old son who goes to bed at an earlier time than my wife and I. We watch movies almost exclusively after he has gone to bed. If my HT were not sound proof, there would not be any time at which I could watch anything that was not PG. At least not at a sound level above about 80 dB. At this level almost no special effect movie is very exciting. In other words, my HT would be, for all practical purposes, useless. I may as well just have a television. There are so many instances where the lack of sound isolation would cause a serious constraint on the usage of the HT that I need not elaborate on them, since I am sure they would be readily apparent to anyone.

Then, there is the acoustic problem of the dead room. This is secondary, but I find it appealing that solving the noise problem usually fixes the acoustical problem at the same time. The two are not really separate issues.

By far the most difficult task in sound isolation for the existing room is the HVAC system. It is virtually impossible to install mufflers in an existing room. My recommendation here is to trace the HVAC ducts and try and find some place where the common metal ducts can be replaced by the soft sided Insulated Flexible Ducts. Replace as much of this duct length as possible. Again, all too often these ducts will be inaccessible or the HT will be placed close to the HVAC unit itself. In this later case, replace as many of the ducts going to the other rooms as possible. In my home, I have all of the ducts sound proofed to every room which has a major effect on sound transmission of all noise throughout the home.

I shudder to think about a HT in a home without forced air AC, for a sealed room will quickly become “stuffy” because of the audience and the electrical equipment. I would hazard a guess that the only solution here would be adding a forced air ventilation system, or accepting the fact that sound isolation is impossible. If the forced air system is added then doing this quietly is the critical requirement. I have no experience with this situation so I won’t comment further.

The next thing that I would do is replace the door to the room with a solid exterior door (or install one if it doesn’t have one), one that seals well all the way around. This is a fairly easy modification with a big payback. The price is highly dependent on the kind of door that you choose. Then, look for other obvious sound leakage paths. The rule here is that if air can go through the hole or crack, then so can sound—un-impeded! Fill these holes with urethane foam, and then blend them in to the surroundings with plaster, paint, whatever.

If your are more ambitious, it is not too difficult to hang an interior wall on an existing wall using the same RC-1 Resilient channel that I used earlier. This will give much of the same effect as constructing the entire wall as I have described. A CLD (Figure 11-1 on pg. 200) construction of drywall can also be added. However, if you are going to do all that much work, then why not just strip out the interior sheet rock and build the walls right—with separate studs for each side of the wall? This is not much more work and the results will be substantially better.

Finally, there is the ceiling—ah, the ceiling, there is never an easy solution to this problem. Of course, there are situations where the ceiling may not be a problem. Hurrah for you, you are indeed lucky—the job's done! But what about the situation where you have an existing room, with living space above? The spaces above will always hear and feel the sound in the HT, and you are just going to have to accept this. You can reduce the sound transmission with a false ceiling, but if it is suspended, then it should be done with nylon cord and not with the steel wires that are standard. For all practical purposes, steel wire is a rigid connection. Nylon cord is incredibly strong and offers a high loss to the transmission of vibrations. The longer the nylon suspension the better, but the room's ceiling height will usually dictate the length, preferably this false ceiling would be heavy and rigid for maximum sound attenuation. Standard drop ceilings of 1/2" (or even 1") ceiling tiles are virtually useless at transmission loss. Basically, this kind of installation will add to the rooms high frequency damping but will do very little for the sound isolation. We're looking for the opposite of these two effects. If the room is lively and needs to be damped, then sound absorbing panels on the ceiling might be a good choice, especially if placed between the listener and the screen. This is because we don't really want refections off of the ceiling as this is detrimental to the sound quality. I usually count on the loudspeakers directivity to reduce this refection, but that is not always the case. Remember, trying to judge if your room is too live or not cannot be done in an empty room. If empty it should always be fairly live.

11.5.b Conclusion

This has been a short chapter when compared to the size of the tasks that this stage of the project encompass—these tasks will take the most time to actually implement. The reason for this is that I have only described the details of the differences between what I recommend and the traditional construction, leaving the details of the traditional construction for outside reading. Someone will have to either know these common construction methods and techniques or will have to read about them or learn them from someone else. But, it is true that someone will have to have this knowledge to construct the room.

When modifying an existing room, I could only give general ideas about how I would approach the problem because there is such a vast array of specific issues and I could not possibly cover them all. I hope that I have given you an idea of what you should be trying to achieve and why, so that you can make more inform judgements about the specific modifications that you are going to implement.

Whatever you do, please don't underestimate the importance of the noise control measures and acoustical treatment which you put into your HT. Simply buying the equipment and installing it in an unmodified room will almost certainly lead to a suboptimal situation. I've been in this kind of room (often) and they always disappoint me. Usually, the owner is completely ignorant of this fact and I always dread being asked what I think of it. They only know what they've seen and heard before and don't have a good base of comparison. But, once they experience an optimally configured room, they are invariably "blown away" and their old system just doesn't seem the same anymore. We've all experienced this effect with audio systems at some point in time. It's even more powerful when the entire movie experience is the comparison.

CHAPTER 12 *INSTALLING THE EQUIPMENT*

To many people, this chapter is likely to be the only chapter of interest to them. Admittedly, equipment selection, purchasing and setup is the most fun. Unfortunately, one needs an appreciation for the fact that the problem is much bigger than just the equipment. I have seen this happen so many times in audio—the belief that all one needs to do is buy the right equipment and it will sound good in any situation. Oh, if it were only that simple.

If you have jumped here through boredom with the rest of book, that's OK, I guess (at least you did buy the book!), but the person who is has arrived here after the diligent reading and comprehension of the rest of the book will be in a much better position to utilize the information in this chapter to create a true Premium Home Theater. It may be arrogant of me to presume this, but I think that a really good home theater could not be achieved without attention to most of what is in the rest of this text.

12.1 System configuration

The first choices in system selection are dictated by how the system is configured. There are many ways to do this. I will develop this aspect of the problem by first showing, in block diagram form, what the equipment has to do. Then, I will describe the different options for implementing these requirements. I probably won't describe all of the options, I may not even know them all, but I will describe the ones that have worked well for me.

Figure 12-1 shows the block diagram of the basic building block of the Audio/Video portion of a HT. I'll discuss these blocks one at a time, cross referencing back to previous discussion whenever applicable.

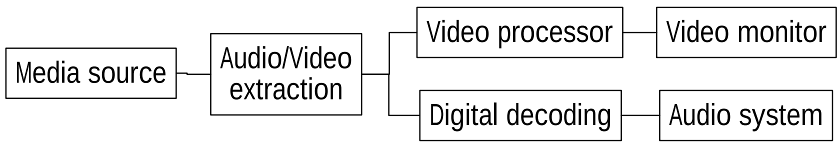


Figure 12-1. Block diagram of HT AV functions that need to be performed.

12.1.a Media Source and Audio/Video Extraction

For all practical purposes the primary video source for a HT will be DVD (see "DVD" on pg. 149.). It could be D-VHS tape (see "D-VHS" on pg. 152.) or, as a last resort, VHS tape, although I won't discuss VHS since I consider it to be obsolete. Another possibility is HDTV (see "HDTV" on pg. 142.). I also won't discuss legacy television in any detail.

DVD is the easiest to implement since DVD players are readily available. Some even contain the digital audio decoding block and some have video processing. DVD video outputs come in two basic varieties, S-video and Component Video. S-Video is the poorest video signal that I would ever recommend using for a large screen. Its problems are readily apparent on a screen much larger than about 5 ft. It is basically a two component system with the chrominance (see "Legacy Formats" on pg. 140.) and luminance signals (see "Luminance" on pg. 123.) on separate lines. Therefore, S-Video is a two value color map which can never be as good as a three color value system. It came about as an improvement to the composite video signal which dates back to the first color television systems. It is basically obsolete by today's standards and is kept around only for compatibility with older systems. By 2006, it will be completely eliminated by FCC decree.

Many, if not most, DVDs today have Component Video outputs. Component Video is derived from the original RGB signal that is obtained by the video camera (see “Three Color Systems” on pg. 124.). Component Video and RGB signals are nearly identical in performance, although the component signal is derived from and must eventually be returned back to an RGB signal. My preference is for RGB throughout since no additional processing is required for this signal. Component video is usually what is found on better television’s—ones denoted as HD.

If a DVD player is placed into a computer, then the video processing takes place inside the computer in software. The computer virtually always puts out an RGB signal. This method of extracting the Audio and Video is so important that I reserved this discussion for a separate section.

The audio must also be striped off of the DVD to be output to the audio processing chain and this is done in many different ways as well. A HT should always have at least 5.1 capability. This capability is licensed from Dolby labs and it is not that common to find this being done in the DVD player itself since the licensing fee’s can make the price of these players noncompetitive. Usually, the 5.1 signal is output as an RF signal either optically or through a phono-plug connection. This signal must then be decoded for 5.1 channel reproduction (the .1 stands for a limited bandwidth subwoofer channel). All DVDs output a stereo signal.

There are also several ways to do the 5.1 decoding. The most common is with a receiver or external decoder box. It is becoming more common to find PC sound cards with this capability.

The simplest form of HT-AV presentation then is a DVD player connected to a Component Video monitor with the sound output routed to a digital capable receiver. To me there are several downsides to this configuration. First, the video (being Component Video) requires a Component Video input on the video monitor or a decoder to RBG. This requirement limits the availability of monitor and projectors and most lower-priced projectors don’t have Component Video. The ones that do have Component Video are usually higher priced as a result. I prefer a different approach which may seem like overkill, at first, but it is actually the most cost effective one in the long run. I will describe that approach next.

12.2 The Home Theater PC

12.2.a HTPC

A simple PC computer can be outfitted with a DVD player and many inexpensive sound cards offer digital inputs and outputs. In this configuration, one is basically using the PC as a simple DVD player, just as in the configuration in the previous section. But this comparison is not as simple as it seems. This is because the tremendous power of the PC can be used to combine virtually everything that one could imagine a HT doing into a single box. It turns out that this approach is extremely cost effective and can often be the lowest cost solution to the entire problem of the HT-AV system.

I will caution the reader about one serious drawback of using a PC and that is that you have to use a PC! If PC's turn you off, aggravate you or you simply cannot get comfortable with them, then the HTPC approach is most definitely not for you. I will get back to your options later. The HTPC market is relatively new and evolving and the components tend to be leading edge technologies that don't always work as well as they should or will. I have been using this approach for nearly four years now and, I find that as time goes on these components are getting better all the time, but they can still be quite temperamental. I am pretty experienced with computers and I sometimes find that I simply cannot get a card or piece of software to work right if at all. If I can't do it, then the average person is in real trouble. But, as I said, the situation is getting better all the time.

The first consideration that I want to delve into has to do with resolution. In a PC, resolution is a pretty clear cut. Its simply a matter of selecting the resolution, in pixels, that you want from a list of resolutions on an applet. Of course, the monitor that you have has to support this resolution, and this is not always obvious. Most desktop monitors can handle 1024x768 without too much trouble, but projectors at this resolution tend to be the higher priced units. Although common, this higher resolution is not always the optimum choice.

Recall the discussion about resolution on pg. 157 and in particular Table 9-1, "Optimum Viewing Distance," on pg. 159. Careful consideration at this point can save a lot of money on resolutions that your not going to be able to see anyway. I'm afraid that it is all too common to buy the highest resolution possible, but this is not the most cost effective approach. At any rate, you should have no problem selecting the resolution that is right for your situation from the information given on those pages.

As I described previously (see "Video Media" on pg. 148.) the resolution of DVD is much lower than what is possible today on even the lowest priced monitors, and much lower than what is available at the upper end of the projector market. The change in resolution that is required for a DVD to play on a higher resolution device is typically called scaling (see Interpolation and Scaling on pg. 150). The DVD data must be interpolated to the resolution of the video system for correct playback. As described, this interpolation is no trivial matter and not something that one gets in a low cost (or even medium cost) DVD player. The PC on the other hand has more than enough computational power to perform these interpolations in a high resolution manner.

Early on in the HTPC market, the processors could not read data from the DVD, do the MPEG decoding and complex video interpolation required with the CPUs of that era. They just ran out of power. As of this writing this situation has changed and these tasks can now all be done on the CPU, usually with power to spare. The first PC DVD players had external processing boards with a dedicated MPEG decoder and the interpolation software. These units work extremely well and, in fact, I still use one in my system even now. They would run on processors of under 1 GHz, which the modern software DVD players cannot do with any quality.

It used to be my opinion that software DVD's were not viable since I could not get any of them to work effectively on my system. I now know that processor speed is everything in this regard since I have seen software DVD players that run fine and look as good as my hardware decoder, but they did required processor speeds of upwards of 1.5 GHz.

Today, the MPEG decoding is actually moving off of the CPU and onto the video card as even more video processing capability is being built into the newer video chips. This situation is in such flux at the current time that any comments that I am likely to make will probably be obsolete by the time that this book is published. One of the big points that I do want to make is that the technical details provided in this text will allow you to be in a position to understand the implications of whatever technology comes forth in the future. The real advantage that technology discussions, like the ones that I have done in this book, have over a simple component recommendation is that the technology changes much more slowly than the products. The technology discussions will be valid and useful for many years to come, but the components will be new virtually every year.

Herein lies one of the real advantages of the HTPC. With it, one can keep on top of this technology by simply upgrading a PC card or a piece of software. One doesn't have to buy a completely new system each time an improvement comes along. Often it is just a matter of downloading a new software release to get added functionality.

Another major feature of the HTPC is its ability to do just about any function. Today, PC cards are available for HDTV, PC-VCR's, DVD players, digital audio decoders, etc. In short, just about the only thing that the PC can't do is actually amplify the audio signal, you'll always have to have external power amps and speakers, and a device to actually display the video—a monitor. In this configuration of Figure 12-1, the equipment block diagram will look like that shown in Figure 12-2. The media source

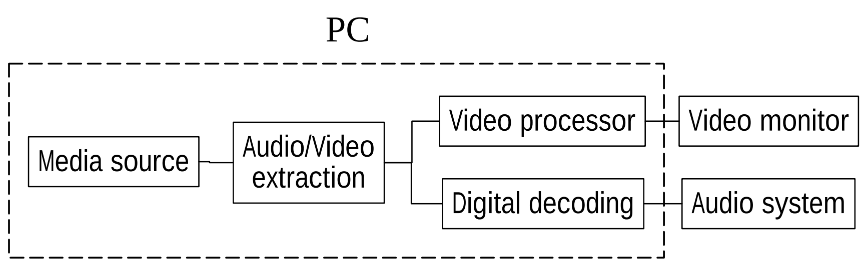


Figure 12-2. A HTPC block diagram.

here is not fixed as being a DVD. It can also be terrestrial HDTV (which at the time of this writing is the only real option for true HDTV), satellite or cable with recording and time shifting capability, or it can even be D-VHS, (since most of these units can connect to a HTPC through IEEE-1394 Firewire). This makes the PC, as a control center, an extremely cost effective solution to the overall Audio-Video control problem.

A new block diagram that shows the layout of the HTPC and the required peripherals in more detail is illustrated in Figure 12-3. There are four important PC cards in this setup and I will talk about each of them in a little detail. I have already discussed the DVD situation and how that can be

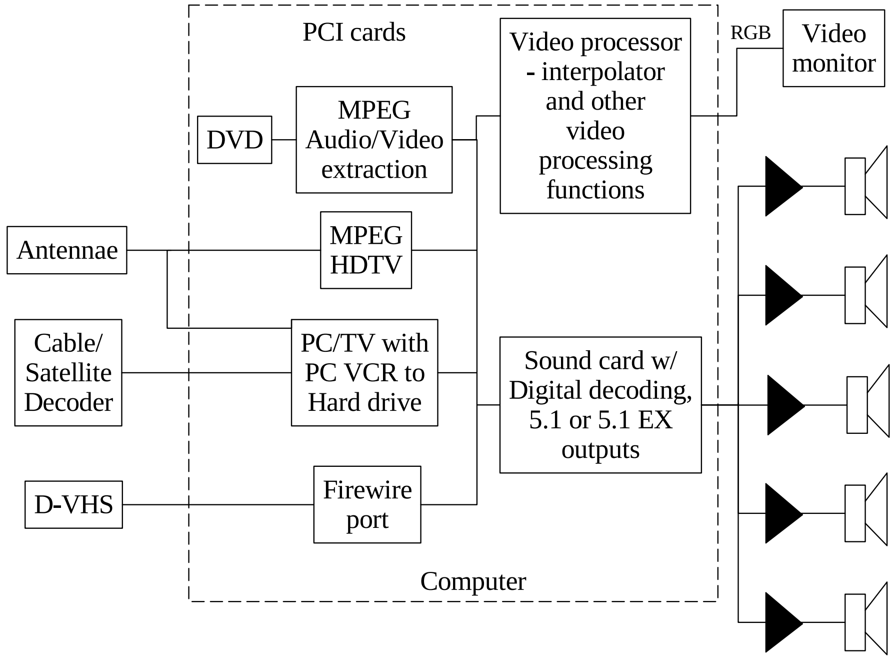


Figure 12-3. Block diagram of a HTPC showing the variety of input formats available as well as the simplicity of the final system configuration. This is by far the most cost effective approach to HT.

handled as hardware or software, although at the present time I do not know of any hardware decoder options. These seem to have all exited the market on the tails of the software DVD players.

12.2.b HDTV PC Cards

At the time of this writing, there are only a few cards on the market which can perform this task. The MDP-100/120 by MIT are the most common cards. I'm not here to make specific component recommendation, just to make you aware of what is available so that you can find the latest and greatest when the time comes. I'd also be remiss to quote prices since these fluctuate so much, but I will say this: the PC based HDTV solution is <u>far less</u> expensive than the external box solution (less than half the cost).

The MDP-100 (with which I have experience) has an S-video input, so it could in fact be used as the input for a satellite or cable connection. Unfortunately, I have not seen a PC card that does the decoding of the raw satellite or cable signals. It is unlikely that this approach will ever be mainstream due to the proprietary nature of these input data streams.

Although, with PC's, anything is possible. Using the S-Video input, this card can play any of the satellite or cable channels, albeit with an external decoder box for proprietary signals. I have not found this card to do a good job of acting as a PC-VCR—for recording off the air programming. It does have an extraordinary capability to record HDTV signals, both to a hard drive as well as to D-VHS via a Firewire connection. These recordings are the best playbacks of a recorded signal that I have ever seen. Unfortunately, the availability of HDTV signals is still rather sparse.

I'll get back to the PC-VCR situation in a moment, but I want to delve into the D-VHS connection a little further. It may be apparent that the availability of video material that is up to the capability of an premium HT is quite limited. DVD's do look very good on these systems, that is until you see a true 1080i or 720p program on this same playback system. Then, the lower resolution of the DVD is readily apparent. There are, at the current time, only two ways that one can get real HDTV (720p or better) in their HT. Terrestrial broadcast, usually only PBS for 1080i or D-VHS, that's about it. There is a lot of talk about HDTV over satellite and cable and this may come to pass, but for me this has not proven to be an appealing option. Perhaps this will change since these distribution channels do seem to be the ideal sources for this program content. But, keep in mind that HDTV takes up almost four or five times as much bandwidth as an MPEG compressed NTSC signal as found on cable or satellite. This means that these service providers must give up a lot of program channels for each HDTV channel that they add. Since their bandwidth is limited, this is a major problem.

I think that it is safe to say that widespread HDTV programming will not become readily available in the very near future and it is hard to tell how fast it will actually arrive. Fortunately, there are, albeit quite limited, sources of these supreme quality video signals available through terrestrial broadcast and D-VHS sources. It may be that these sources will become so established that they become the next standard. It is a long road to create and establish a new format that consumers will accept and this does not seem to be "just around the corner" for HD DVD, although it is a hot topic. While this debate is going on the D-VHS market is growing. If it grows to a sufficient size then it could well become an established standard in video distribution, but the current trends do not appear to support that situation. There are actually fewer and fewer companies making D-VHS machines today than in the past. I have been reluctant to purchase one of these machines because of its uncertain future (I don't think that it would look too good next to my Beta machine).

Transport stream the actual bit stream format for HDTV as it is broadcast.

To communicate with a PC, a D-VHS machine needs firewire and a processor to decode the **transport stream**. Transport stream is the technical name for the data stream that composes HDTV signals. If this stream is recorded to a digital medium, then it can be played back through the MPEG decoding scheme to a full HDTV signal. This is the mechanism by which D-VHS transfers the AV signals to and from the computer. It is also the same data steam that the HDTV card uses to record HDTV signals to disk as data files, which can be played back through the HDTV card at some later time. This works flawlessly for HDTV signals, but cannot be used for analog signals coming into the HDTV card. This is the reason that this card is lacking in a viable method for recording cable and satellite signals.

12.2.c PC-VCR

The PC-VCR is basically a computer that saves video files to the hard disk. Years ago, this was not a viable thing to do because of the size of these files, but today, with hundred Gig hard drives and higher, this is a viable option to removable media. With a DVD writer, the "better" videos can be easily archived.

My experience with these programs is mixed. They are usually quite inexpensive and come with a television tuner card. This card is kind of wasted since we already have an HDTV which can, of course, also receive analog signals. If only the HDTV cards had good recording capability for analog signals, then there would be no need for a PC-VCR purchase. One can only hope.

The quality of one of my PC-VCR's was very good, but its stability was poor. It worked for me for a short time and a friend could never get his to work. The other board that I tried was stable, but I could not get a good recording out of it. I believe that the difference was in the MPEG encoding algorithm. One was using the standard windows codec, which is not too good, while the other was a custom codec, which worked good when it worked. As I said, the HTPC is prone to be difficult to implement, but it tends to work well when it does.

These PC-VCR's work much like any of the modern genre of Hard Drive Video Recorders. They have time shifting, multi program selection and start times. The one major problem that I have seen is when using an external decoding box—the software cannot change the channel of the incoming signal. This has to be done manually in all of the implementations that I have seen. I am sure that in time this too will be possible.

12.2.d Gaming

I am not really a big "gamer", although I do enjoy them from time to time. I will say one thing—once you have played a game on a 12' screen with surround sound, a PC monitor will forever be a disappointment. I have a race car game which is so real that your palms actually sweat during the race. Those people who have experienced this big screen thrill have been impressed, although it's hard to describe.

The important thing is that with a HTPC the gaming capability comes for free, so why not enjoy it!

12.3 Audio

12.3.a Main Setup

The audio hookup of a HT is really pretty straight forward, especially if a PC is used because then there are only six audio outputs, one for each channel. The selection of the speakers was discussed in Loudspeaker Selection on pg. 74, but I have not touched on the placement of the speakers which is an important consideration.

From Chapter 4, I showed that it is better to have a higher directivity source in the more lively rooms that I recommend. With a directional source one can minimize the interaction of the speakers with nearby walls, thereby minimizing the early reflections and the associated coloration. This situation is shown in Figure 12-4. The three main speakers are assumed to have a narrow directivity of nominally 90° in the horizontal plane. The directivity coverage is shown as the dotted lines in this figure. The left speaker is shown facing forward and the right is canted in. The canted

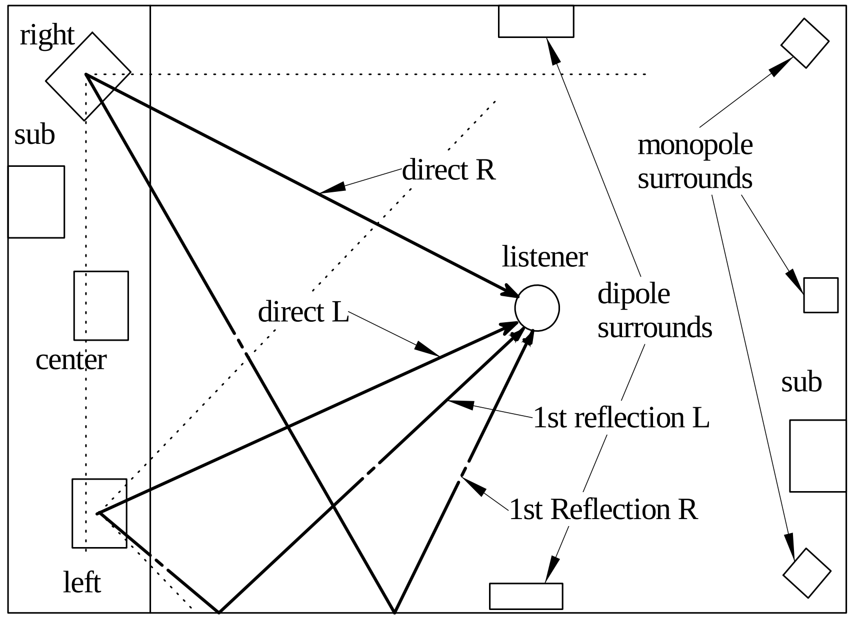


Figure 12-4. Loudspeaker layout for home theater with 6.1

speaker is preferred in this room and the left speaker is only shown facing forward for demonstration purposes.

The reason that canting the speaker is advantageous can be seen by looking at the direct and reflected waves. The direct sound from the left speaker arrives principally from the left and is followed very shortly by a wall reflection which also arrives from the left. The right speaker, on the other hand, does not have much energy impinging on the nearest wall due to its directivity and so the direct sound arrives from the right speaker in much the same way as the left; however, the first reflection is considerably more delayed, and, just as importantly, it arrives from the left. The configuration shown for the right speaker has a couple of advantages. First, the arrival of the direct and early reflections are from opposite sides of the room—to opposite ears. Then there is the extended delay of the first reflection due to its longer path length. Both of these help to minimize the coloration due to comb filtering. For omni-directional loudspeakers, there is no advantage to be gained for any direction of the enclosure. Also, a negative condition for omni-directional loudspeakers is that there are also very early reflections bounced off the back wall, in addition to the side walls. Clearly,

*Figure 12-5.
Dipole surrounds shown hanging from nylon cord for low vibration transmission. The door and a stone wall are visible in the background.*

the directivity of the source has a major effect on the direct to early reflection ratios in a small room.

Also shown in Figure 12-4 are the locations of the surround speakers. Placing the surrounds is almost always heavily influenced by the room layout, door placement, equipment cabinet, etc., but I have shown some suggested placements. The side surrounds are normally dipoles. I use two smaller speakers (smaller than the main speakers), usually two ways, which are placed back to back and wired out of phase. This yields a dipole with the dipole axis pointed toward the seating line. In this way, even though the surround speakers are nearby the listener, the sound from them is always diffuse owing to the lack of a direct field along the dipole axis.

Figure 12-5 shows a side surround dipole loudspeaker made up of two smaller two-ways glued back to back. They are suspended from nylon cord to reduce sound and vibration transmission up to the structure. As I have said, nylon cord is very effective at this. The cords were then painted black, resulting in an invisible support.

Dolby Digital EX
a newer format where a sixth channel is encoded through a difference signal to the left and right surrounds.

In most theaters, the surround speakers are smaller than the main speakers, but there are usually many of them. Likewise, in a HT, it would be desirable to have many surround speakers. I would add one to the back wall, which can also be used as a sixth channel in **Dolby Digital EX**. This I would make a dipole. I might also add a pair of surround monopoles to the rear corners. The monopoles placed like this will help to offset the power low frequency response of the dipoles at the sides.

In my experience it is important to have the front three loudspeakers of identical design. This prevents an image wander that can occur when these three speakers have different frequency responses—which will virtually always be the case if they are not identical in design. This situation is shown in Figure 12-6. This photo shows the commercial theater speakers, and a back wall covering which reduced both a refection off of the back wall from the speakers as well as light from the projector (the screen was porous). Making the surrounds identical is not as critical, but it should be done if possible.

Figure 12-6. Front loudspeaker array, left, center and right shown with the screen lifted out of the way. Note black absorbing material on back wall for both sound and light.

12.3.b The Subwoofer

LFE
Low Frequency Effects—a band limited channel on the movie sound track that is exclusively for frequencies below 100 Hz.

There is a whole mystique surrounding subwoofers that I won't get into. This is the least critical of all of the speakers due to its limited bandwidth. In the main three frontal speakers the woofers will almost certainly be large because that is the only way to get high directivity. As such, they will provide a great deal of the low end for the HT. But, there is a **Low Frequency Effects** (LFE) channel in Dolby Digital and the presence of this channel virtually requires the use of a subwoofer.

As I said, I do not consider the design of the subwoofer to be critical, unless it is poor or over-driven. Then, it could produce distortion that is quite audible and annoying. Ported or closed is pretty much the same to me as long as they are done right. Today, it is so "cookbook" to design and build a subwoofer that they are sold by the millions at very low prices. Do not buy a high end subwoofer, it is not as critical as the full range front three channels.

The solution that I always recommend, which is not always doable, is to buy two or three subwoofers that are not too small, but they need not be the biggest either. About 12"–15" is a good choice. Then I would place these subwoofers as follows: the first one should go into any corner; the second one should go near, but not ”in”, the opposite corner; finally, if you do add a third, then place it along a long wall near its mid point. The exact locations of the subwoofers are not critical, but the more that there are the smoother the low frequency response is likely to be.

12.3.c Wiring

Another major subject of great mystery is how to wire the audio system. In a cable, there are only two elements that are important, the connector and the wire. The connector should be robust since it can often be tugged, pulled, twisted, and bent and standing up to this abuse without failing is an important criteria. There is also the coating, which can corrode over time yielding a less than desirable connection. Gold does work well here, but only if it is thick enough to not scratch off in use—which most are not. To purchase gold coating is expensive and not always worth the money. But, beyond these two characteristics, there is little else to consider in a connector.

The next aspect is the wire. Apart from its construction, shielded or not, there are two features of importance in a wire. The first is its capacitance and the second is its current carrying capability. A high capacitance wire as a high level (not speaker cables) connection will reduce the high frequency response. The longer the wire the worse this situation gets and the more important the capacitance issue becomes. Short interconnect cable of less than six feet can use almost any shielded wire. Longer runs, like to other rooms, will benefit by a low capacitance cable—usually thicker.

Construction enters into the equation in various ways. For high level signals, the cable must be shielded with a grounded shield, otherwise it can pick up hum from the lighting etc. The gap between the shield and the conductor is where the capacitance comes from. The greater this gap the lower the capacitance (material is a smaller factor). For really critical applications, the connections are sometimes a pair of twisted wires with a grounded shield. These constructions usually require special connections at each end to work most effectively and are not usually applicable to most HT components, although they are the standard for pro gear.

Finally, there is the current capability. For interconnects of high level signals (low current) this aspect is irrelevant, but for speakers it is somewhat important. Thicker cable here is always better, up to a point. It is also not necessary to shield the speaker cables because of their high current (low impedance level). But things like Oxygen Free Copper and massive cable sizes are simply a waste of money. Just go to the corner hardware and buy 10 or 12 gauge copper zip (or lamp) cord. The speaker cables can usually be routed anywhere and length is not an issue if they are thick enough.

12.3.d Calibration

Audio system calibration can have a significant effect on the sound system's performance. But, if not done correctly, like any calibration, can seriously degrade the system's performance. Audio calibration is no exception.

There are two major aspects of the audio system that needs to be calibrated. The first is the channel levels and the second is the channels frequency response. The first is straightforward to do and most systems provide signals for doing this, but the second is extremely difficult, requiring substantial knowledge about what you are doing. It can also require a great deal of specialized and often expensive equipment. A review of

Appendix IV - Acoustic Measurements will give an indication about how to perform these tests.

Setting the levels is simply a matter of listening to a test tone sent around to the various speakers and adjusting the level controls until the sound level is equal. This is usually easy to do, but can be somewhat tricky when the speakers are all different and have different tonal characteristics. It is hard to judge what are tonal differences from what are level differences. This is just a natural occurrence for the human hearing system. But don't worry, the overall performance of the sound is not heavily dependent of the level adjustments.

Setting the LFE channel level is different. That's because it is not usually fed with a test signal in the setup options. My suggestion is to not set this channel too high because two hours of a loud LFE channel will get on your nerves. There is also not set standard here that I have found. By that I mean that different movies and music signals already have a vast array of low frequency content and levels. It is not uncommon to have to adjust the LFE from source to source to get an acceptable mix.

EQ
Equalization, the technique of correcting a sound systems frequency response electronically.

Calibrating the frequency response is also called **equalization** or EQ. First, to do EQ, you have to have an equalizer. In many HTPCs, there is an equalizer built into the software. Otherwise, you must have an external piece of electronic equipment. These devices are available in a wide variety of costs and capabilities. There are two basic types, octave or one-third octave and parametric. Parametric is my recommendation, but they are the rarest and usually the most expensive as a separate component. But no matter what type of equalizer you use, there is always the question of how to set it. Professionals do not trust their ears and use a variety of expensive measurement equipment and techniques to calculate the optimum settings. This procedure is usually beyond the capabilities of most HT constructors and is a topic in and of itself. If you are not well trained or don't have the right equipment, then I don't recommend doing the EQ this way. Be leery of a so-called professionals who do have the right equipment but may not know how to use it correctly. If they are not using the techniques that I describe in this book, like spatial averaging, don't trust them.

As a novice, you will have to trust your ears, but this is notoriously inaccurate. I recommend setting the controls by ear with several listeners joining in and judging the EQ with various source signals. There are CD's for just this purpose. Things like noise and impulses can be most revealing, but don't over EQ. The adjustments should always be as small as possible—try and avoid "painting a skyline" with the EQ settings. But, properly done, listening adjustments can lead to a much better sounding system.

12.4 *Video*

12.4.a Wiring

Projectors are often located many feet from the source and the scaling/ processor components, so getting the video signal to the projector intact is paramount. Most projectors are connected to the source or processor by an RGB or component video connection. These signals operate at very high frequencies and are analog. This makes the cabling and connectors a primary consideration. Unlike audio, where the cables and speaker leads are pretty much irrelevant, the quality of the wire (its capacitance and inductance, etc.) are important factors in image quality. A loss of high frequencies in these signals can mean a significant loss of image quality. The cables should be as short as possible. Making custom cables can be tricky-because the construction of the cable, particularly attaching the connectors, has to be done right or signal refections and degradations will occur. Here, it is best to buy pre-made cables of high quality, unless you are confident that you can assemble high quality ones.

DVI
Digital Video Interface—an all digital interface where there is no digital to analog and analog back to digital conversion required for a digital based monitor.

There are also emerging direct digital interfaces, the most prominent of which is DVI, or **Digital Video Interface**. Using the DVI interface with a compatible source allows the video signal to remain in the digital domain until the projector, this can eliminate many of the issues associated with analog cables. Other emerging digital interfaces include IEEE1394 (Firewire) and HDMI. These are not widely implemented as of this writing.

Color reference
the temperature that corresponds to a chosen color of gray for the recording and playback systems.

12.4.b Calibration

In addition to proper room and projector set-up and screen size and type selection, a complete system requires calibration. Movies and television programs are mastered to a set standard color of gray as a **color reference**, namely 6500K. In addition to gray scale, user controls such as contrast, brightness and color saturation and tint must also be set correctly.

To do this calibration yourself requires a DVD made expressly for this purpose. Having or borrowing one of these DVD's is essential. They provide the test patterns and instructions for their use. Since I recommend getting one of these discs which includes instructions, I won't repeat these instruction here. I will mention that on the one that I use, I found the set-up instructions and recommendations for the audio quite disappointing, so I do not recommend following the guidelines in this regard. Instead, rely on the vastly more complete recommendations made in this book.

12.5 Installing the projector and screen

Most of the aspects of the projector installation have been covered already—keep it out of the room if possible, check on throw length, etc. Figure 12-7 shows the projector mounted behind a plate glass sound barrier.

Figure 12-7. Projector mounted behind the rear wall and extending through an angled glass plate.

Note that the glass is at an angle to the lens. This plate glass is a special non-reflective, high transmission glass available from an optical parts supplier, like Edmund Scientific. The projector is in another sealed room behind the HT's rear wall. This room houses all of the equipment and some of the HVAC mufflers.

Installing the screen is not very difficult, but it must be tight and flat so that there is no curvature in the screen surface. I used copper tubing for the screen frame in mine. It is smaller than the PVC plumbing shown in Figure 12-6. The scren can be sewn on or you can use Velcrow. Note that there is a center support to keep the frame from bowing. This design worked well allowing the screen to be hinged up out of the way for access to the speakers behind it. There are so many other options that one can employ to mount a screen that I won't go into any more detail. The screen fabric handles much like any fabric and can be dealt with exactly the same.

12.6 Conclusion

At this point, I wish that I had a grand send off to encourage you to go out and build your own HT. If you have read this whole book, then by now you know as much as many professional installers and more than many. You are in a position to direct the project or do it yourself with full confidence that the end result will be something to be proud of and enjoy.

As an encouragement, I will leave you with some photos of my own HT as it currently stands. This will highlight some features that I used and recommend and some that I don't. It will also allow you to see a different design than the previously documented HT. These photos are shown on the next page.

My HT uses 3/4" oak plywood panels suspended by nylon cord for the ceiling, which worked quite well. The floor is oak, which has been a bit of a problem over concrete due to weather changes. The screen is perforated with the center channel behind. This HT has a little over $10,000 in material costs (at today's prices)—much of that the projector. The labor was all my own. My family and I enjoy it immensely.

Figure 12-8.
The authors HT looking towards the screen. This is a 4:3 setup 1024x768, DLP projector. Note cultured stone wall on right and (light) sound absorbing wall quilt on the left. ALso on the left is the game wheel for a driving simulator. The speakers are behind the black grills. Plans are to convert this setup to 16:9.

Figure 12-9.
The authors HT looking from the screen towards the door. Note the equipment cabinet and surround speaker. This cabinet gets in the way and I do not recommend this approach. I plan to move it into the right portion of the rear wall. Note the sparse furnishing and leather seating. The HTPC is in the cabinet and the projector can be seen on the ceiling. The subwoofer is behind my son's head

APPENDIX I - Subjectivism

I would bet that everyone has experienced the kind of complications that can result from the widely different subjective opinions that we all have. In this regard, there are some very serious issues surrounding the subjective assessment of HTs, audio systems, video systems, etc. that I would like to discuss. The label that I use to describe these issues is **subjectivism**.

Because of the fundamental subjective nature of the subject matter dealt with in this text, Audio-Visual (AV) playback systems, I feel obliged to give some words of caution and some advice on this subject. It is important for people to come to terms with subjectivism since it is the basis of an enormous problem in the marketplace. Marketing people love the fact that AV has a personal subjective impression as its final criteria, because it is well know that this impression can be easily manipulated. This manipulation occurs, to a great extent, because of the complex technology behind AV where it is often difficult for the lay person to sort out the facts from the opinions—the wheat from the chaff.

As an example of what I am talking about I will relay one of my favorite stories about subjectivism. Some years back a large car company was studying the "sound quality" of various cars. In an experiment, they showed subjects photos of a car while playing a sound recording of it through a high quality playback system. The subjects were then asked to evaluate the "sound quality" of these presentations. In another part of the experiment, the sound and images were randomized so that the two things were no longer in sync. Much to their dismay, the only correlation which could be shown after the randomization was between the rating and the picture, the actual sound appeared to be irrelevant, even though it was the sound that the subjects were asked to evaluate! That's how strong the impact of brand recognition, or marketing pitches, etc. can have on subjective assessments. A good brand can do no wrong; if only that were actually true!

Another example actually occurred to me on the video side of subjectivism, which is not really so different from the audio side. I was once in a hi-end HT store and I noticed a $6000 price tag on a projection screen. Curious at how they could justified this price tag, since it was nearly ten times the cost of an equivalent sized screen from a more reasonable supplier, I asked the salesman about this. The salesmen responded, with some indignation that I would even ask such an foolish question, that the “color rendition of this screen was the only one acceptable in a premium Home Theater system”. The color rendition? Excuse me, but the screen was white just like any other screen. There is no “color rendition” difference between two white screens. There are some aspects of the screen that make visible differences (see Screens on pg. 174), but these factors are well understood and quantified and the two screens (the high priced one and mine) had the exact same values for both of these measures—color and gain. I suspect that the extra $5400 was for the highly acclaimed brand name which, of course, the salesman was so enthusiastic about. His commission was based on this higher selling price.

To further add fuel to this fire, while I was waiting for my “cheap and inferior” screen to arrive, and being as impatient as anyone else to see my projector in use, I went down to a white goods sale and bought a bed sheet for $10 to use as a temporary screen. When the professional screen arrived it was found to be noticeably, but not substantially, better than the bed sheet screen and to some people (like my wife) it was not even appreciably different. In her opinion it certainly did not justify a $590.00 price difference, let alone a $5990.00 one! The point here is that I am not a big fan of components whose price is not justified by its performance. **Value** is the word that I like to use, I want value for my money, not hype.

It is with great relief that over the years I have come to see many of Hi-Fi’s myths fall by the wayside. Discussion of things such as speaker wire size and oxygen content, or CD edge color, are mostly gone now—thank goodness, but, unfortunately, it is becoming clear that “Home Theater” has become the new frontier for exaggerated marketing hype. My hope is that the technical discussions in this book will help to dispel these myths—facts tend to do that. Having the facts at ones disposal naturally leads to a better foundation for reliable opinions.

When dealing with subjectively oriented AV issues, I prefer to deal with the science (facts), whenever possible, and not **Audio Tarot**. My wife coined this wonderful phase after she first came into contact with the audio community that I work in. She has a background in experimental psychology and knows how easily subjective opinions can be swayed by external factors. She noted that a large amount of audio folklore is accepted by its practitioners purely on faith, i.e. there is no way to either prove or disprove these beliefs. Basically, audio, in these aspects, is a religion. Mankind has always had trouble reconciling facts with their fundamental belief system when the two come into conflict (to wit Gallileo). The facts are often suppressed in order to perpetuate the established belief system. The practitioners of Audio Tarot will always prefer to suppress those facts that contradict their established beliefs.

It is not possible to deal with a subjective subject such as AV without resorting to some subjectivism, although much of the time a purely scientific approach will go a long way. Still, it is usually necessary to let some subjectivism into the discussion. For example, in my screen example there is simply no room for opinion, the science is well established. However, there is room for a subjective opinion about the use of a grey screen versus a white one, but not when the two screens differ only by a brand name.

Deriving data (facts) from subjective impressions is an interesting challenge and one that can be done sucessfully, but it is a tremendous amount of work to do it properly. This is certainly the reason why so little scientific work has actually been done on the subjective aspects of AV. This is more true of the audio portion of the problem than the video one since a great deal of scientific work has been done by the television industry on the perception of video. Video preceded audio and even today it remains the more well studied of the two. Bell Telephone did do a substantial amount of work on the perception of sound in the early half of the last century, but mostly from the communications standpoint. Much of our understanding of hearing today came from that work. Unfortunately, the audio world, where music quality is the goal, lags far behind the study of the perception of communication signals (voice).

For the most part, subjective assessments in audio are not scientifically valid because they lacked sufficient experimental controls—double blind, sufficient subjects, etc. There are researchers, like Floyd Toole, who do an outstanding job at this task, but they are the exception, not the rule. It is still all too common to see people trying to make sound quality assessments with biased, uncontrolled subjective listening tests.

In this text, I attempt to give the reader enough theoretical background in the underlying subjects of AV to be able to make an informed judgement of both the validity and the importance of sales information that they might get about products that they are considering purchasing. It is quite common for a corporate marketing group to attempt to confuse the consumer with data that falls into these two categories. The first, validity, refers to product claims that are based on data which is not very accurate or quite possibly incorrect. The second, importance or relevance, refers to product claims that, while they may be perfectly correct, are simply not relevant. For example, one can hardly find an amplifier without specifications quoting outstanding THD and/or IMD numbers. Unfortunately, these numbers tell you nothing about how the amplifier actually sounds—the data is valid but it is simply not relevant. This same amplifier might also be described as "clean" or "transparent" without any scientific meaning being given to these subjective terms. These descriptors therefore lack validity. Valid and relevant performance data is sometimes hard to find.

Hopefully the reader of this text will begin to be more comfortable with the ins and outs of subjectivism and learn when to question the claims, when to ignore them, and when to pay attention to them. This is not altogether an easy assessment to make, even for the professional.

Appendix II - *Logarithms*

II.1 .. The basic Logarithm function

The mathematical function which relates the linear and dB scales is known as the logarithm and is written as $log(x)$. The shape of the log curve is very important and a plot of $log(x)$ for values of x from .001 to 100 is shown in Figure II-1 on pg. 248 on two different plot types—a linear x axis on the top and a log x axis on the bottom. The log of 1 is zero and number less than one are always negative, smaller values being more and more negative leading to the log of 0 as negative infinity. The $log(10)$ is one, and the $log(100)$ is 2 and so forth. The important feature of the log function is that the larger x is, the slower $log(x)$ changes with x. If $log(x)$ is plotted with a logarithmic x axis, then the curve is simply a straight line. This is one of the key features that are of interest, since the perception of a physical sensation tends to be logarithmic in nature. Thus using the log function on physical values (like sound pressure or light intensity) will give a linear (straight line) relationship between the perceived level and the actual physical level. This relationship is extremely important.

An important implication of log values is that two numbers multiplied together will simply add if they are both in the log domain. For instance multiplying a log value by two is accomplished by simply adding .301 to the value (.301 being the log of 2). Multiplying by 10 simply adds 1.0 to the value. Division works likewise with subtraction—dividing by 10 simply subtracts one from the original value. Another important implication of log values is that a log value raised to a power simply changes the slope of the relationship on a log plot. In the bottom graph of Figure II-1, the dashed line is a plot of the log of x^2. Each successive power simply increases the slope of the line. In filter specifications, the order of the filter is synony-

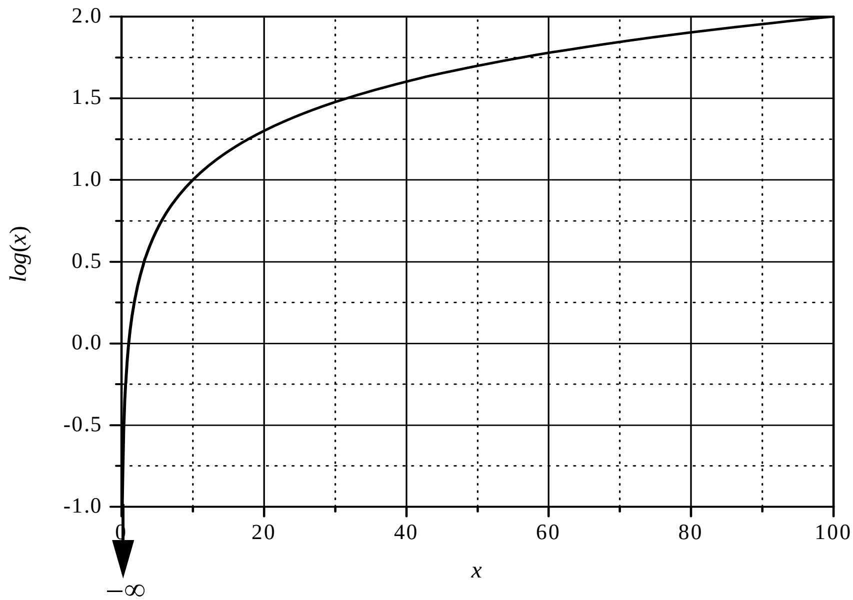
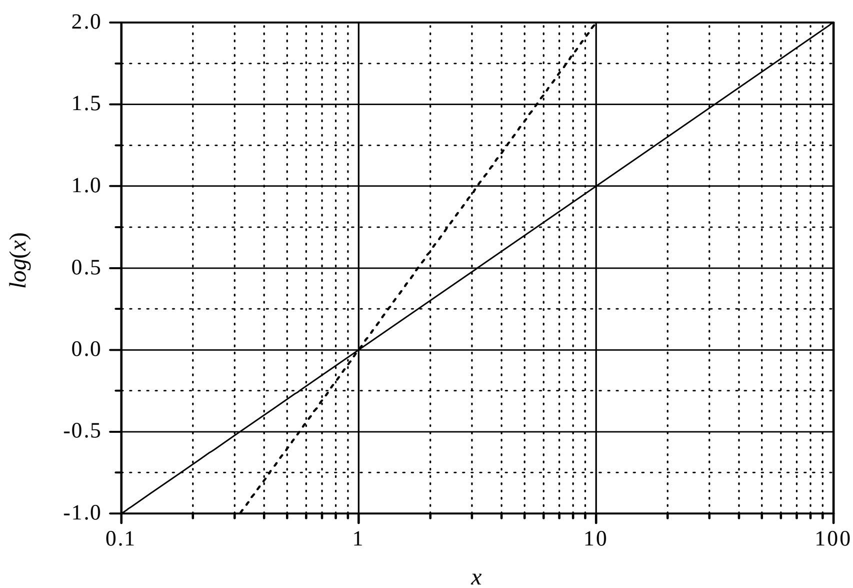


Figure II-1.
The logarithm function - linear scale (top) - log scale (bottom)



mous with this logarithmic power relationship—a higher order filter is a higher power in frequency.

With the log function under our belts I can now define the dB scale, which is our real interest. A deci-Bel (dB) is the ratio of two powers and is mathematically defined as

$$dB = 10 * \log\left(\frac{power_1}{power_2}\right)$$

Note that there are no units associated with the dB scale, it is dimensionless. Generally the quantities that we will deal with are not powers, but power is usually proportional to the square of the physical quantity, like Watts and Voltage. This means that the dB definition for non-power quantities (the only ones that I use in this text) like pressure or Volts is

$$dB_{pressure} = 10 * \log\left(\frac{pressure_1^2}{pressure_2^2}\right) = 20 * \log\left(\frac{pressure_1}{pressure_2}\right)$$

where I have used a mathematical rule of logarithms that states,

$$\log\left(x^n\right) = n\log(x)$$

(which is the same feature that I talked about above). This law states that raising the argument (inside the parentheses) of a log to a power is equivalent to multiplying the log of the argument result by this same number. Since dB is always a power ratio, the underlying values must always be **RMS** (see pg. 9) values, not peak values. RMS is kind of an average value taken over time. The difference between the peak value and the RMS value is worth noting, but not worth worrying about. For voltage levels in audio equipment, the dB reference is usually 1 volt, although **VU** (Volume Units) references the maximum level before clipping in the circuitry.

VU
Volume Units—the level of a signal relative to the maximum level capability of the system.

SIL
Sound Intensity Level—the sound power per unit area.

We must always remember the power ratio requirement. For example, **Sound Intensity Level** (SIL) is a power so its dB scale uses $10\log(x)$ not $20\log(x)$.

In log scaling, each doubling of a non-power quantity results in an increase of 6dB in level, halving it results in –6dB. Note that adding two identical signals is simple in the dB system. Since two signals have twice the value, this results in a $20\log(2) = 6$dB increase in the level. Four signals would be 12 dB and a ten signal increase the level by 20 dB. There is a small caveat in this addition, and that is that the signals must all be identical. Adding dissimilar signal levels is not so easy.

The dB scale is extremely important for even the most basic understand of audio and video. Fortunately, both fields use the same definition of dB as their basic quantity of measure. In noise control, it is important to note that a 6 dB reduction in noise is equivalent to cutting the physical Sound Pressure Level in half. A 20 dB reduction is an enormous 90% reduction. When viewed in these terms, it is easy to see why noise isolation of a room, where 40–50 dB reductions might be desired, is such an arduous task.

Appendix III - *Complex numbers*

I expect that the introduction of the mathematical concept of complex numbers in Chapter 1 may have intimidated many readers. As much as I have attempted to make this text free of “complex” math, some concepts are simply fundamental. Actually, understanding the basic definition of a complex numbers is not really that difficult, so I have done it here.

Figure III-1 shows a complex number as the point at the tip of the arrow in what is called the complex plane. Complex numbers have several features which make them useful in the study of waves (audio, radio, light), which is, that they can be used to represent wave motion in a very concise manner. Complex numbers inherently incorporate the magnitude and phase aspects of waves that I talked about in Chapter 1.

Phasor
a vector in the complex plane which denotes the location of a complex number.

The arrow—lets call it a **phasor** (just like on Star Trek) since that’s what mathematicians call it—can be described in two ways, by its magnitude (the length) and its phase (the angle away from the horizontal), or by its

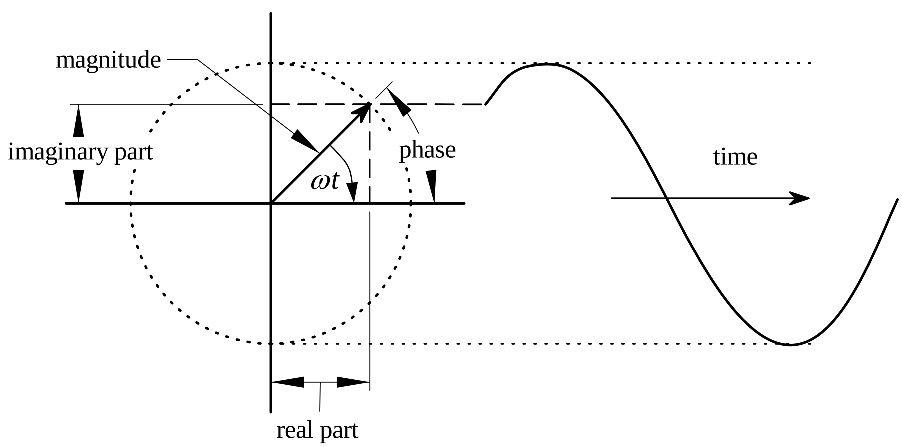


Figure III-1. Complex number representation of a frequency data point.

horizontal and vertical components. The phase can vary from 0 to 360 degrees, or more commonly, from 180 to –180 degrees. The horizontal and vertical parts are called the real and imaginary parts respectively. Note that the magnitude and the real and imaginary parts form a right triangle.

Consider the phasor being rotated around the origin at the rate of f times (cycles) per second (Hz). It will sweep out an angle of $360*f*t$ degrees ($2\pi*f*t$ in radians), which is the phase angle that I talked about. It is convenient to use $\omega = 2\pi f$ where the frequency variable ω is now in radians per second. The difference between f and ω is a simple constant 2π, but using ω instead of f saves a great deal of writing (a lot of 2πs) in many situations.

As the vector rotates around the origin at the rate of ωt, the imaginary part of this phasor conveniently traces out the time waveform of a sine wave as shown on the right hand side of the Figure III-1. The circle comprises all of the points that the waveform goes through as time progresses, repeating itself every $1/f$ seconds. The most important aspect of a complex number is its magnitude (the length of the line which we can also see is the amplitude of the sine wave), but we must always remember that any discussion of waveforms must be cognizant of the fact that signals in the frequency domain always have both a magnitude and a phase.

Think of Figure III-1 as being a single dimensional display of a three dimensional space, with the third dimension, frequency, going into and out of the page (we would then have many parallel pages). The frequency domain representation of a signal is then comprised of all of these planes with each one rotating at its specific frequency. The total time signal is then simply the sum of the values of the imaginary parts of each plane at any instance in time. This may seem complicated, and it is, but envisioning it will allow you to see the equivalence of the time and frequency domains—they are nothing more than two different ways of looking at the same thing.

It is actually conventional to use the real part of the phasor as the time domain signal, but that did not lend itself to the drawing as shown, so the wave is traced out by the imaginary part instead. Normally one takes the real part of an imaginary number to get the instantaneous time value of the waveform.

APPENDIX IV - Acoustic Measurements

IV.1 . Analog Measurements

Analog measurements form the oldest and easiest to understand of all measurements. Unfortunately, they are rather limited and cannot do nearly as much as the more modern digital techniques can do.

SLM
Sound Level Meter—an analog meter for measure the broadband sound level.

The simplest analog instrument for sound is a **Sound Level Meter** (SLM). There are some very inexpensive SLMs available in the marketplace, but these days a sound card works just as well, perhaps not as well calibrated in absolute level. These meters often have a weighting scale selector (see Figure 6-1 on pg. 100). When used carefully, one can get some useful information from these devices. They have two main problems. First, they measure all frequencies at the same time (perhaps weighted) and second, they measure all of the sound (noise, refections, etc.). Within these constraints, they can be used to do some rudimentary equalization and channel level matching.

Another fundamental instrument is a Volt Meter, sometimes called a Volt-Ohm-Meter or VOM. These usually measure AC and DC voltage and current levels. The AC measurement is intended to be the RMS value of the signal at the input, with some time averaging. For complex signals most inexpensive meters are not “true RMS”, they will have errors in the level. In the past, a strip chart recorder was used. This device consists of a long strip of paper that would be moved under a pen that moves perpendicular to the papers motion. The location of the pen is made to correspond to the level (usually in dB) of a measured signal. If a narrow band filter and signal generator are synchronized to the sweep rate of the paper then this strip chart will record the frequency response of the Device Under Test (DUT) as a frequency seep is sent to the DUT. This type of measurement is sometimes seen even today, although the technique is obsolete. Very similar techniques are available in software programs.

IV.2 ..Digital Measurements

Record length the length of time that data is taken.

Number of samples the total number of individual samples in the record.

Digital measurements are by far the most common ones in use today. In these digital measurement systems, there are two primary parameters that influence the data that we get when we sample a signal. The first parameter is the sample rate that was introduced in Chapter 1 and the second is the **record length**, or the **number of samples** in the record. The record length is measured in seconds and the number of samples is an integer. In principle the number of samples could be very large, but in practice it is usually a number based on a power of two (for simplicity in calculating the FFT), such as 512, 1024, 2048, etc. Notice that the sample rate and the number of samples determines the record length and visa versa, from the equation,

$$record\ length = sample\ rate \cdot number\ of\ samples$$

A very large number of samples will require a very large record length—the measurement will take a lot of time.

I could simply reduce the sample rate and thus allow for a large number of samples in a short period of time, but this approach has other problems. Short time records will yield poor resolution of the frequency bins. This is easy to explain. As the sample rate goes higher, the "valid" frequency range (below the Nyquist frequency, see pg. 138) gets larger. This frequency range is covered by the same number of frequency bins, so each bin must therefore get wider. Mathematically this is described as follows,

$$\Delta f = \frac{f_s}{N}$$

Δf = the frequency bin width

f_s = the sampling frequency

N = the number of samples

The frequency bin width also determines the lowest frequency of data. For example, taking 1024 data points sampled at 44.1kHz (common numbers for these parameters) will give us a low frequency resolution of 43 Hz, which is a poor resolution at the low end. There would only be two data points below 100 Hz. Stepping up this resolution to 4.3 Hz would require that we either sample at 4.41 kHz or extend the number of samples from 1024 to 10240. This later requirement would raise the record length from 23 ms to 230 ms. A clean record of this length is difficult to obtain and store.

When we measure a loudspeaker in a room, the sound goes in all directions, bounces off of the walls, and some of it ends up at the measurement microphone. There are some complex situations which result from this fact. If I use a steady state tone then I will simply get the power response of the loudspeaker at that frequency—as modified by the room. There is also the problem of extraneous noise in this measurement. I could use noise and some slick signal processing tricks (cross-correlation) to get the response at many frequencies with the extraneous noise rejected, but it would still be the room modified power response—just cleaner. Getting the direct response is more complicated.

The room reflections are delayed in time by the excess distance that they have traveled. If I were to use data that goes directly from the source to the microphone, the direct sound, then I would need to take all of the data before the arrival of the first reflection. From the above discussion we can see that getting a 4.3 Hz resolution from 20–20 kHz requires a reflection free data record of 230 ms. If the sound travels at the speed of 344 m/s then this means that the excess distance that the reflected sound must travel needs to be,

$$344\ {}^{m}\!/_{s}\ \cdot .23\ s \cong 80\ m$$

With the source and microphone placed at the center of the room, this equates to a room no smaller than 40 m in any direction from the source and microphone pair. This is not a practical thing to do indoors—that's a <u>big room</u>. Many loudspeaker companies actually do there measurements out of doors where they can afford this kind of space, but, of course, they must deal with the environmental noise, which, as I said, can be processed out of the data. An anechoic chamber alleviates this problem to some extent, but very few of them are truly anechoic below about 100 Hz. And then they cost a couple of million dollars too.

Gating
using only a portion of the impulse response of the system.

MLS
Maximum Length Sequence—a pseudo-random noise with special properties.

Window
a weighting function applied to a time signals to focus in on a particular aspect of the data.

When on wants to measure only the direct sound from the loudspeaker in a real room, then **gating** must be used. One can use many techniques (noise, **MLS**, swept sine, etc.) to calculating the impulse response of the system (see pg. 15), but this response will contain the room reflections. By applying a **window** the data, we can gate out the initial, refection free portion of the impulse—the direct response. This gating technique works quite well at higher frequencies, but if the room is small then the window must be short and the low frequency resolution will be poor.

Another common technique is to place the microphone very close to the source and simply ignore the reflections, since they will be so much lower in amplitude than the large "nearfield" pressure signal. This technique works well for improved low frequency resolution, but leads to significant problems of nearfield standing waves as the frequency of analysis is increased. The best that one can do is to blend the nearfield measurement with a gated far field measurement, but this can be tricky to do—especially with multiple woofers or ported systems. The point here is that it is not reasonable to do full-range, simulated anechoic measurements in rooms of reasonable sizes. We will need to either accept lower resolution, which will be most evident at the lower frequencies or a limited upper frequency range—in order to focus on the lower frequencies. Or we could simply allow the presence of the reflections (as is sometimes done, although, interpretation here is a problem). There is simply no good way around this problem.

What is the importance of the direct versus steady state field measurement? Consider EQing a loudspeaker in a room. If one does a steady state measurement of the speaker in the room then basically they are measuring the power response of the system (modified slightly by the room). As an example I will use the system of Figure 4-5 on pg. 69. I know that this system is basically flat on-axis (from the line along 0°). However, from Figure 4-6 on pg. 70, I also know that the power response is, most certainly, not flat. If I measure the loudspeaker response in steady state, I am going to get a measurement analogous to the curves shown in Figure 4-6. If I equalize this measurement to flat then I will have a distinctly non-flat axial and direct response. So you can see that there is a marked difference in EQing to the direct versus the steady state measurement—unless the system is Constant Directivity (see pg. 71), which is what I always recommend.

INDEX

M

N

O

P

T

V

W